AI Horisontti



Hiljainen *suostumus*

Tietoturvatutkijoiden mukaan koodausavustajan lisäosan uuteen versioon oli upotettu takaovi. Käyttäjän mukaan pois kytketty tietosuojavalinta palasi päälle kenenkään sitä valitsematta.

TALOUS
Teknologiajättien piilevä tekoälyvelka kasvoi puolessa vuodessa 90 %.

TIETOTURVA
Kylmälaitehyökkäykset aiheuttivat vahinkoja myös Suomessa.

TYÖKALUT
Sanelu ja viestien lajittelu toimivat jo omalla koneella.

PÄÄKIRJOITUS · VIIKKOKATSAUS

Älä kysy enää

Kesken ajatuksen ruudulle ponnahtaa kysymys: saako avustaja ajaa saman komennon kuin äsken? Kiireinen ihminen valitsee vaihtoehdon, jonka jälkeen samaa ei enää kysytä, koska jokainen kysymys katkaisee työn juuri siinä kohdassa, jossa se oli alkanut sujua. Valinta on hiljainen suostumus, joka säästää yhden klikkauksen joka kerta. Viikon tietoturvauutiset kulkivat lähes kaikki samanlaisten mukavuuksien kautta.

Tietoturvatutkijoiden mukaan koodausavustajan lisäosasta levisi versio, johon oli upotettu takaovi eli hyökkääjälle jätetty piilotettu sisäänkäynti. Altistuminen ei vaatinut käyttäjältä virhettä, koska lisäosat päivittyvät usein itsestään. Yrityskäyttäjä kertoi, että pois kytketty valinta, joka sallii keskustelujen käytön mallien kouluttamiseen, oli OpenAI:n ohjelmointiavustajassa palannut päälle kenenkään sitä valitsematta. Toinen ohjelmointityökalu lähetti vikailmoituksen mukaan käyttäjän sähköpostiosoitteen sivustoille, joilta se haki tietoa. Saman työkalun nappi, jolla lupaa ei kysytä uudelleen, liitettiin tietoturvayhtiön kirjoituksen otsikossa salasanoihin 1 122 julkisessa tiedostossa. Kaupoissa sama kaava toimi ilman tekoälyä: Kyberturvallisuuskeskuksen mukaan hyökkääjät pääsivät käsiksi etäohjattaviin kylmälaitteisiin ja muuttivat lämpötiloja niin, että ruokaa pilaantui. Vahinkoja syntyi myös Suomessa. Osio ”Mitä tämä tarkoittaa käytännössä” seuraa kuuden hengen tilitoimistoa viikon läpi ja kertoo, mitä kehittäjältä ja huoltoyhtiöltä kannattaa kysyä.

Viikon toinen kysyjä oli raha. Nikkei Asian mukaan yhdysvaltalaisten teknologiajättien piilevä tekoälyvelka kasvoi kuudessa kuukaudessa 90 % ja on nyt 3,2 biljoonaa dollaria. Torstaina teknologiaosakkeita painottava Nasdaq-indeksi laski 5 %, kun sijoittajat Wall Street Journalin mukaan vetäytyivät tekoälyyn sidotuista osakkeista. Vastapainoksi OpenAI julkaisi asiakastarinoita, joissa säästöt kuvataan sujuvasti mutta ilman tarkkoja lukuja. Osio ”Markkina, sääntely ja turvallisuus” kertoo, miksi suurten tekoälyyhtiöiden pörssilistautuminen toisi näkyviin, kuinka kestävällä pohjalla nykyiset hinnat ovat.

Samaan aikaan osa tekoälystä muutti pilvestä omalle koneelle. Kehittäjien suosikkilistoille nousi sanelu- ja puheentunnistustyökaluja, joilla palaverin muistio syntyy ilman, että äänitallenne lähtee talosta. Työkalujen suomen taito selviää vasta omalla puheella kokeilemalla. Vapaasti ladattavan mallin turvarajat voi kuitenkin poistaa kuka tahansa kotikoneellaan. Mallien jakelupalvelun suosituimpien joukossa oli viikolla tällainen riisuttu versio yli 70 000 latauksella. Osio ”Työkalut ja avustajat arjessa” kertoo, mitä omalla koneella voi jo tehdä ja mitä kannattaa välttää. Tekninen osa menee syvemmälle ja näyttää muun muassa, miten saastuneen lisäosan jäljet etsitään omalta koneelta.

Sähköposti- ja salasanatapaukset tuskin jäävät viimeisiksi, joten työkalujen päivitystiedoista kannattaa jatkossa lukea kohdat, joissa mainitaan henkilötiedot, tunnukset ja tallennetut luvat. Kerran annettu lupa ei viikon perusteella pysy paikallaan. Kone lakkasi kysymästä, joten kysyminen jäi ihmiselle.

TÄSSÄ NUMEROSSA

KAIKILLE
Mitä tämä tarkoittaa käytännössä
Työkalut ja avustajat arjessa
Markkina, sääntely ja turvallisuus
Mitä seurata

TEKNINEN OSA
Suuret toimijat
Avoimet mallit ja ekosysteemi
Ajoalustat ja infra
Benchmarkit ja vertailut
Työskentelytavat ja agentit
Liitännäiset, MCP ja repot
Tietoturva ja haavoittuvuudet

VIIKON KÄRKI

Viikon tärkein asia ei ollut uusi tekoälymalli. Tärkeintä oli toistuva havainto siitä, että tekoälyavustajalle kerran annettu luottamus ei pysy paikallaan ja tämä painaa sitä enemmän, mitä laajemmin avustaja pääsee käsiksi käyttäjänsä tiedostoihin ja tunnuksiin. Maanantain ja sunnuntain välillä nähtiin tietoturvatutkijoiden mukaan takaovi koodausavustajan automaattisessa päivityksessä, käyttäjien raportoima itsestään päälle palannut tietosuojavalinta ja ohjelmointityökalu, jonka kautta raporttien mukaan sähköpostiosoitteita päätyi verkkosivustoille ja salasanoja julkisiin koodivarastoihin. Kuvaa muutti torstai, jona teknologiaosakkeita painottava Nasdaq-indeksi laski Wall Street Journalin mukaan 5 % sijoittajien vetäytyessä tekoälyyn sidotuista osakkeista. Päivästä toiseen jatkuivat kauppojen kylmälaitteisiin kohdistuneet hyökkäykset ja tekoälyn siirtyminen omalle koneelle, kun taas yhtiöiden itse laatimat mittaristot, luvuttomat asiakastarinat ja turvarajoista riisuttujen mallien latausluvut osoittautuivat lähinnä kohinaksi.

---

1 **Takaovi tuli päivityksen mukana.** Anthropicin koodausavustajan lisäosa levisi maanantaina Cursor-yhtiön koodieditorin lisäosakaupassa versiona, johon tietoturvatutkijoiden mukaan oli upotettu takaovi eli hyökkääjälle jätetty piilotettu sisäänkäynti. Altistuminen ei vaatinut käyttäjältä virhettä, koska lisäosat päivittyvät usein itsestään. Yrityksen, jonka kehittäjät tai alihankkijat koodaavat tekoälyavustajan kanssa, kannattaa kysyä heiltä asiasta suoraan.

2 **Avustaja kertoi käyttäjästään liikaa.** Perjantaina kehittäjien keskustelupalsta Hacker Newsissa lähes 500 kommenttia kerännyt vikailmoitus kertoi Anthropicin ohjelmointityökalun lähettäneen käyttäjän sähköpostiosoitteen sivustoille, joilta se haki tietoa. Viikonloppuna tietoturvayhtiö Apiiro kytki kirjoituksensa otsikossa saman työkalun mukavuusvalinnan, jolla samaa lupaa ei kysytä uudelleen, salasanoihin, jotka ovat päätyneet 1 122 julkiseen tiedostoon koodinjakopalvelu GitHubissa. Kummassakin tapauksessa tieto kulki reittiä, jota käyttäjä tuskin olisi itse valinnut.

3 **Kerran tehty tietosuojavalinta ei pysynyt.** Tiistaina yrityskäyttäjä kertoi kehittäjien palstalla, että OpenAI:n ohjelmointiavustajan asetus, joka sallii keskustelujen käytön mallien kouluttamiseen, oli kytkeytynyt takaisin päälle kenenkään sitä valitsematta. Kyse on käyttäjän havainnosta, jolle OpenAI:n selitystä ei viikon lähteistä löydy. Asetuksen tilan näkee ChatGPT:n tietoasetuksista hetkessä.

4 **Kaapattu tekoälytili tulee kalliiksi minuuteissa.** Tivin mukaan rikolliset kaappaavat yritysten tekoälypalveluja ja käyttävät niitä toisten laskuun. Yhdessä tapauksessa kaapatulle tilille kertyi minuuteissa 100 000 dollarin lasku. Käytön mukaan maksavan yrityksen kannattaa tarkistaa, onko tilillä kulukatto ja kuka reagoi hälytykseen myös iltaisin.

5 **Kylmälaitehyökkäykset ulottuivat Suomeen.** Kyberturvallisuuskeskuksen mukaan hyökkääjät ovat eri puolilla Eurooppaa päässeet käsiksi kauppojen etäohjattaviin kylmälaitteisiin ja muuttaneet niiden lämpötiloja niin, että ruokaa on pilaantunut. Tiistaina Suomessa oli havaittu epäiltyjä tapauksia ja perjantaina vahinkoja oli syntynyt myös täällä, vaikka suurimmilta vältyttiin. Kylmälaitteita käyttävän yrityksen kannattaa selvittää, voiko laitteita ohjata etänä ja kuka sitä yhteyttä hallitsee.

6 **Raha alkoi kysyä tuottoa.** Nikkei Asian mukaan yhdysvaltalaisten teknologiajättien piilevä tekoälyvelka kasvoi kuudessa kuukaudessa 90 % ja on nyt 3,2 biljoonaa dollaria. Torstaina Nasdaq-indeksi laski 5 %, kun sijoittajat Wall Street Journalin mukaan vetäytyivät tekoälyyn sidotuista osakkeista. Tekoälypalveluja käyttävälle yrittäjälle luvut kertovat, millaisten panosten varaan nykyiset hinnat on rakennettu.

7 **Tekoäly asettui omalle koneelle.** Saksalainen Nextcloud, jonka ohjelmistolla organisaatiot jakavat tiedostoja omalla palvelimellaan, valitsi tekoälytoimintoihinsa täysin avoimen eli koulutusaineistoaan myöten julkisen mallin, joka toimii tavallisella kotikoneella. Omalla koneella toimivat sanelu-, puheentunnistus- ja viestienlajittelutyökalut nousivat viikon mittaan kehittäjien suosikkilistoille. Luottamuksellista aineistoa käsittelevälle yritykselle tämä tarkoittaa, että asiakirjat voivat pysyä talossa ilman, että tekoälystä luovutaan.

01 Mitä tämä tarkoittaa käytännössä

Kuuden hengen tilitoimisto on antanut asiakasportaalinsa ylläpidon freelancer-kehittäjälle, joka tekee työnsä kahden eri yhtiön tekoälyavustajan kanssa. Maanantaiaamuna hänen koodieditorinsa oli yön aikana päivittänyt toisen avustajan lisäosan uusimpaan versioon kysymättä mitään. Tiistaina hän olisi voinut huomata, että keväällä pois kytketty asetus, jonka piti estää keskustelujen käyttö mallien kouluttamiseen, oli jälleen päällä toisessa avustajassa. Perjantaina avustaja haki verkosta ohjetta maksupalvelun liittämiseen ja saattoi samalla kertoa sivustolle hänen sähköpostiosoitteensa.

Sunnuntaina hän muisti painaneensa joskus kiireessä nappia, jonka jälkeen avustaja ei enää kysynyt lupaa samaan komentoon, vaikka komennossa oli tietokannan salasana.

Hän ei asentanut viikon aikana mitään epäilyttävää eikä klikannut yhtäkään kalastelulinkkiä. Jokainen neljästä reitistä kulki jonkin mukavuuden kautta: automaattisen päivityksen, itsestään palanneen asetuksen, avustajan avuliaisuuden ja napin, joka säästää yhden klikkauksen joka kerta. Tilitoimiston kannalta kyse ei ole kehittäjän huolimattomuudesta vaan siitä, että asiakkaiden tiedot

kulkevat työkalujen läpi, joiden asetuksia kukaan talossa ei näe.

Maanantain takaovessa luottamus annettiin kerran asennushetkellä ja uusittiin sen jälkeen hiljaa jokaisen päivityksen mukana. Anthropicin koodausavustajan lisäosa levisi Cursorin lisäosakaupassa eli sovelluskaupan kaltaisessa jakelupaikassa versiona, johon tietoturvatutkijoiden mukaan oli upotettu haittaohjelma, joka antaa hyökkääjälle etäyhteyden koneeseen. Koodausavustajan lisäosa on houkutteleva kohde, koska sen tehtävä on lukea koodia, ajaa komentoja ja käsitellä samoja tiedostoja, joissa on usein myös salasanoja ja pääsyavaimia eli ohjelmien välisiä kirjautumistunnisteita. Julkisista otsikoista ei voinut päätellä, tapahtuiko murto Anthropicin omissa järjestelmissä vai lisäosakaupan julkaisuketjussa. Saastuneet versiot ja tunnistetiedot löytyvät Cursorin tietoturvatiedotteesta ja tapauksen ensimmäisenä raportoineen tietoturvayhtiö Phorionin kirjoituksesta.

Yrittäjän ei tarvitse ymmärtää teknisiä yksityiskohtia. Riittää, että hän kysyy kehittäjiltään ja alihankkijoiltaan, onko lisäosaa käytetty Cursorissa tai muussa samaa lisäosakauppaa käyttävässä editorissa ja onko se päivittynyt tapauksen aikaan. Myönteinen vastaus tarkoittaa, että konetta kannattaa kohdella murrettuna. Silloin vaihdetaan salasanat ja mitätöidään pääsyavaimet, joilla kirjaudutaan esimerkiksi pilvipalveluun tai maksujärjestelmään. Samalla pilvitileiltä tarkistetaan, onko niillä näkynyt epätavallista toimintaa.

Perjantain sähköpostitapaus kulki toista reittiä. Jo tiistaina Anthropicin ohjelmointityökalun päivitystiedoissa kerrottiin, että työkalu kysyy jatkossa luvan, ennen kuin verkkoa käyttävä tai komentoja suorittava toiminto lähettää käyttäjän tilin sähköpostiosoitteen verkkoon. Perjantaina asiasta tuli Hacker Newsin puheenaihe, kun vikailmoitus, jonka mukaan työkalu lähettää osoitteen sivustoille, keräsi lähes 500 kommenttia. Alle kaksi tuntia keskustelun alkamisen jälkeen yhtiön koodivarastoon lisättiin suoja. Se estää komennot, joissa osoite kirjoitettaisiin ohjelman tunnistetietoon eli siihen tekstiin, jonka ohjelma kertoo itsestään jokaisen verkkopyynnön mukana.

Suojan rakenne vihjaa, mistä on kyse. Osa verkkopalveluista pyytää automaattisia hakijoita jättämään yhteystietonsa juuri tähän kohtaan. Avulias tekoäly voi noudattaa pyyntöä ja käyttää ainoaa osoitetta, jonka se tietää, eli käyttäjän omaa. Tapauksen laajuutta ei viikon aikana kerrottu. Sähköpostiosoite on silti EU:n tietosuoja-asetuksen tarkoittama henkilötieto. Siksi tekoälyavustajia käyttävän yrityksen kannattaa ensin päivittää työkalut ja sitten selvittää, millä tunnuksilla ne on kirjattu sisään. Monelle pienelle yritykselle toimiva käytäntö on erillinen tunnus tekoälytyökaluja varten. Silloin vuotava tieto on vähemmän henkilökohtaista ja sen kulun voi jäljittää.

Viikonlopun löytö koskee mukavuutta, jota lähes jokainen tekoälyavustajan käyttäjä on joskus käyttänyt. Ennen komennon ajamista avustaja kysyy luvan ja käyttäjä voi vastata, ettei samaa tarvitse kysyä enää uudelleen. Tietoturvayhtiö Apiiron kirjoituksen otsikon mukaan tämä valinta on Anthropicin ohjelmointityökalussa johtanut salasanojen päätymiseen 1 122 julkiseen tiedostoon GitHubissa. Hacker Newsissa aihe keräsi viikonloppuna satoja kommentteja. Otsikko ei yksin kerro, miten salasanat tiedostoihin päätyivät. Yksi mahdollinen reitti on se, että pysyvä lupa tallentuu tiedostoon komentoineen. Komentoon näkyviin kirjoitettu salasana kulkisi silloin mukana, kun projekti jaetaan julkisesti. Kehittäjältä kannattaa siksi kysyä kaksi asiaa: kirjoitetaanko salasanoja suoraan komentoihin ja päätyvätkö tekoälytyökalujen asetustiedostot julkisiin koodivarastoihin. Myönteinen vastaus jompaankumpaan tarkoittaa, että salasanat vaihdetaan.

Tiistain havainto OpenAI:n asetuksesta näyttää saman ongelman toisesta suunnasta. Hacker Newsiin kirjoittanut yrityskäyttäjä kertoi, että hänen työpaikallaan pelätään hänen sanojensa mukaan vainoharhaisuuteen asti yrityksen tiedon vuotamista suurille tekoälyyhtiöille. Siksi tietojen jakaminen oli kytketty pois kaikista ohjelmointiavustajista. Maanantaina hän ja hänen kollegansa huomasivat, että OpenAI:n avustajassa asetus oli palannut päälle kenenkään heistä sitä valitsematta. Kirjoittajan mukaan samaa on raportoitu palstalla aiemminkin. Johtuuko ilmiö ohjelmavirheestä, asetusten uudistuksesta vai jostain muusta, jää auki, koska OpenAI:n selitystä ei viikon lähteistä löydy.

ChatGPT:n tietoasetusten kohta "Improve the model for everyone" ratkaisee, saako OpenAI käyttää keskusteluja mallien kehittämiseen. Kerran tehty valinta kannattaa käydä läpi uudelleen suurten ohjelmapäivitysten jälkeen. Sama koskee muidenkin toimittajien avustajia, joihin työpaikalla syötetään luottamuksellista tietoa. Kestävämpi suoja on sopimus. Yritys- ja tiimitilien ehdot poikkeavat yleensä henkilökohtaisten tilien ehdoista, jolloin koulutuskäytön kielto ei riipu yksittäisen työntekijän valintaruudusta.

Kiireinen ihminen painaa nappia, joka lupaa, ettei samaa kysytä enää, koska kysymys katkaisee ajatuksen kesken työn. Avustajat on rakennettu tekemään työstä sujuvaa ja juuri sujuvuuden kautta viikon vuodot kulkivat.

Sama kaava toistui viikon toisessa turvallisuusuutisessa, jossa tekoälyä ei ollut lainkaan. Kauppojen kylmäaltaat ja kylmähuoneet ovat nykyään usein verkossa, jotta huoltoyhtiö voi seurata ja säätää niitä etänä. Kyberturvallisuuskeskuksen mukaan hyökkääjät pääsivät eri puolilla Eurooppaa käsiksi tällaisiin laitteisiin ja muuttivat niiden lämpötila- ja sulatusasetuksia niin, että elintarvikkeita pilaantui. Suomessa vahinkoja syntyi, vaikka keskuksen mukaan suurimmilta vältyttiin. Kauppiaan, ravintoloitsijan tai kiinteistön omistajan kannattaa kysyä huoltoyhtiöltään kolme asiaa: onko laitteessa etäyhteys, onko sen oletussalasana vaihdettu ja tuleeko laitteeseen päivityksiä. Mukavuus, joka säästää käynnin takahuoneeseen, avaa oven myös niille, jotka eivät koskaan käy paikan päällä.

Viikon toinen käytännön opetus koskee laskua, joka kasvaa käytön mukaan. Torstaina Tivi kertoi hakkereista, jotka kaappasivat yhtiön tekoälytilin ja kerryttivät sille minuuteissa 100 000 dollarin laskun. Käytön mukaan laskuttava palvelu ei tunne iltaa eikä viikonloppua: lasku kasvaa niin nopeasti kuin koneet ehtivät lähettää pyyntöjä. Hyvä salasana ei riitä tilin suojaksi, koska ohjelmat käyttävät palvelua usein rajapinta-avaimella eli salasanan kaltaisella tunnuksella, joka voi unohtua koodiin, jaettuun tiedostoon tai tekoälytyökalun asetustiedostoon.

Tarkistus kannattaa aloittaa selvittämällä, ketkä ja mitkä ohjelmat käyttävät yrityksen avaimia. Tarpeettomat avaimet poistetaan ja ne, joiden kulkureittiä ei tunneta, vaihdetaan. Kulukatto tai hälytys kannattaa asettaa aina, kun palvelu sen sallii. Hälytykselle tarvitaan myös nimetty ihminen, joka reagoi siihen iltaisin ja viikonloppuisin. Anthropic toi torstaina Clauden yritysasiakkaille kokeiluun rajauksen, jolla rajapinta-avaimia voi käyttää vain organisaation hyväksymistä verkko-osoitteista, mikä vähentää vuotaneen avaimen arvoa varkaalle. Samaan aikaan tullut uloskirjaus kaikilta laitteilta ei sen sijaan yksin riitä lähtevän työntekijän tai kaapatun tunnuksen hoitamiseen, koska jäsen pysyy organisaatiossa ja voi kirjautua heti takaisin. Vastaavia rajauksia kannattaa etsiä myös muiden toimittajien hallintanäkymistä.

Myös kieli näkyy laskulla. Tivin mukaan tekoälyn käyttö suomeksi voi maksaa selvästi enemmän kuin käyttö toisella kielellä. Käytön mukaan laskuttavat palvelut eivät laske tehtäviä vaan tokeneita eli pieniä tekstinpaloja, joina ne lukevat tekstiä. Suomen pitkät taivutetut ja yhdistetyt sanat pilkkoutuvat yleensä useammiksi paloiksi kuin englannin lyhyet sanat. Tilitoimisto, joka ajaa joka yö 100 suomenkielistä ostolaskua tekoälypalvelun läpi, maksaa siis myös kielestä. Kiinteällä kuukausimaksulla ero ei näy laskulla. Käytön mukaan maksavan yrityksen kannattaa ajaa sama työ kerran suomeksi ja kerran englanniksi ja katsoa palvelun käyttöraportista, kuinka suuri ero omassa työssä on, ennen kuin automaatio ehtii kasvaa.

Tivin mukaan Revolut ratkaisi oman kuluongelmansa toisella tavalla. Yhtiön koodareiden tekoälykulut karkasivat käsistä, kunnes siirtyminen avoimiin kielimalleihin pudotti ne murto-osaan. Avoimen mallin tiedostot on julkaistu vapaasti, minkä ansiosta yritys voi ajaa mallia omalla tai vuokratulla palvelimella ilman käyttökertoihin perustuvaa laskua. Mallin voi myös ajaa eurooppalaisessa konesalissa, jolloin asiakastieto pysyy EU:n alueella. Uutisesta ei käy ilmi, mitä malleja Revolut käyttää tai mihin tehtäviin. Kysymys on silti sama Revolutin kokoiselle yhtiölle ja kolmen hengen markkinointitoimistolle: tarvitaanko kalleinta mallia jokaiseen tehtävään vai riittäisikö osaan kevyempi vaihtoehto?

02 Työkalut ja avustajat arjessa

Viikon työkalumuutoksissa toistui sama suunta: itsenäisesti tehtäviä hoitavat tekoälyavustajat eli agentit siirtyvät erillisistä kehittäjätyökaluista tavalliseen keskusteluikkunaan ja saavat samalla äänen ja muistin. Omalle palvelimelle asennettava suosittu keskustelusovellus on monelle organisaatiolle tapa käyttää eri yhtiöiden malleja yhdestä näkymästä. Sen sunnuntaina julkaistussa versiossa mallivalikkoon ilmestyvät Anthropicin ja OpenAI:n koodausagentit, jos sovellukseen on kytketty agentteja tarjoava komentoympäristö. Agentille voi kirjoittaa tai puhua kuin mille tahansa mallille ja sen pyynnöt käyttää työkaluja hyväksytään suoraan keskustelussa. Kukin keskustelu käyttää yhtä agenttia eikä agentin vastausta voi luoda uudelleen tai muokata jälkikäteen. Jo tiistaina sama sovellus oppi näyttämään ja ohjaamaan siihen kytketyn tietokoneen työpöytää, jolloin tekoäly voi käyttää ohjelmia niin kuin ihminen hiirellä ja näppäimistöllä. Tekoäly toimii silloin koneella samoilla oikeuksilla kuin käyttäjätunnus, jolla se on kirjautunut, minkä vuoksi ominaisuutta kannattaa kokeilla erillisellä koneella, jolta ei ole pääsyä tärkeisiin tiedostoihin. Keskusteluikkunassa annettava hyväksyntä on samaa lupaa, jonka pysyväksi tekeminen liittyi viikonlopun salasanalöytöön.

Ohjelmointieditori Cursor lisäsi lauantaina agentteihinsa äänikeskustelun. Muutostoiveen voi kertoa ääneen ja katsoa samalla ruudulla olevaa sovellusta, mikä on luontevampaa kuin vian kuvaaminen kirjoittamalla. Ennen käyttöönottoa kannattaa selvittää, ymmärtääkö toiminto suomea ja missä puhe käsitellään, koska avokonttorissa mikrofoni poimii helposti myös työtovereiden keskustelut. OpenAI puolestaan kokosi ohjelmointiagenttinsa tekstikomennoilla käytettävään työkaluun näkymän, joka näyttää usean samanaikaisen tehtävän tilan ja merkitsee lukemattomat tulokset samaan tapaan kuin sähköposti.

Anthropic julkaisi lauantaina kirjoituksen siitä, mistä osista agentin muisti rakentuu. Avustaja, joka muistaa aiemmat keskustelut, asiakkaan toiveet ja keskeneräisen projektin vaiheet, säästää aikaa, koska taustaa ei tarvitse selittää joka kerta uudelleen. Sama ominaisuus tarkoittaa, että käyttäjän joskus kertoma tieto säilyy palveluntarjoajan järjestelmässä ja voi nousta esiin aivan toisessa yhteydessä. Käytössä olevan avustajan asetuksista kannattaa tarkistaa, onko muisti päällä, mitä sinne on kertynyt ja miten sen voi tyhjentää. Asiakastietoja käsittelevän yrityksen on EU:n tietosuoja-asetuksen mukaan pystyttävä kertomaan, minne asiakkaiden tiedot päätyvät, myös silloin, kun ne on kirjoitettu muistavalle avustajalle.

Kehittäjien keskustelupalstalla esiteltiin torstaina ilmainen työkalu, joka kirjaa omalle koneelle, mitä koodausagentti teki ja millä perusteella. Kyse on varhaisen vaiheen projektista. Suunta on silti sama kuin toimittajien omissa päivityksissä: agentin tekemisistä halutaan jälki, jonka käyttäjä voi itse lukea.

Toinen viikon kehityskulku vei päinvastaiseen suuntaan, pilvestä omalle koneelle. Maanantaina GitHubin nousevien projektien listalle ilmestyi kaksi ilmaista sanelusovellusta, jotka muuttavat puheen tekstiksi kokonaan omalla tietokoneella. Käyttäjä painaa pikanäppäintä ja puhuu, minkä jälkeen teksti ilmestyy siihen ohjelmaan, jossa kursori on. Perjantaina samaan joukkoon liittyi Microsoftin avoin puheentunnistus, joka litteroi puhetta reaaliajassa ja erottelee, kuka sanoi mitäkin. Siitä on pakattu versio, joka toimii tavallisella suorittimella ilman erillistä näytönohjainta. Palaverin muistio syntyy silloin ilman, että äänitallenne poistuu omalta koneelta. Lääkärin vastaanotolla, asianajotoimistossa tai asiakaspalvelussa ero on olennainen, koska pilvipalveluun lähetetty sanelu on tietoa, joka on luovutettu toiselle yritykselle.

Suomen kieli jää kaikissa kolmessa avoimeksi kysymykseksi. Sanelusovellusten kuvauksissa ei kerrota, miten ne tunnistavat suomea. Microsoftin työkalussa kieliä on kymmenen eikä esittelyotteesta käy ilmi, kuuluuko suomi niihin. Laatu kannattaa kokeilla omalla puheella, ennen kuin sanelutyö siirretään koneelle. Ohjelma ladataan vain projektin omalta sivulta eikä kopioista tai hakukoneen mainoslinkeistä, koska viikon ensimmäinen uutinen näytti, mitä saastunut jakelu voi tehdä.

Pienyrittäjälle konkreettisin uutuus oli perjantaina julkaistu päätösmallien perhe, jonka voi ajaa omalla koneella ilmaisella ohjelmalla. Päätösmalli lukee tekstin ja vastaa siitä esitettyyn kysymykseen valitsemalla vaihtoehdon, arvioimalla kyllä- ja ei-vastausten todennäköisyyden tai antamalla arvosanan. Julkistuksen esimerkissä malli kertoo, onko kyse laskutusasiasta, kun asiakas kirjoittaa, että hänen kortiltaan on veloitettu kahdesti. Asiakasviestien lajittelu onnistuu näin ilman, että viestit lähtevät pilvipalveluun. Mallin kaksi suurempaa kokoa osaa lukea myös kuvia, esimerkiksi kuitin tai vauriokuvan. Esimerkissä käytetty versio toimii toistaiseksi vain Applen omilla suorittimilla varustetuissa Maceissa eikä luokittelun tarkkuudesta ole riippumattomia tuloksia. Kokeilu kannattaa tehdä viesteillä, jotka on jo käsitelty käsin, jolloin koneen vastauksia voi verrata omiin. Samaan ohjelmaan tuli myös Xiaomin pieni malli, joka on tehty ohjelmointiin ja agenttien tehtäviin ja joka toimii tavallisella kannettavalla.

Nextcloudin valinta tiistaina vei saman ajatuksen organisaatioiden tasolle. Yhtiö otti tekoälytoimintoihinsa yhdysvaltalaisen tutkimuslaitoksen Ai2:n mallin, jonka koodi ja koulutusaineisto ovat julkisia mallin itsensä lisäksi. Ai2:n mukaan malli toimii tavallisella kuluttajatason laitteistolla. Organisaatio, joka jo säilyttää tiedostonsa omalla palvelimellaan, voi siis käyttää tekoälyä niin, että asiakirjat eivät kulje ulkomaisen pilvipalvelun kautta. Kirjoitus on mallin tekijän oma eikä siitä voi päätellä, miten malli pärjää suurille pilvimalleille suomenkielisissä tehtävissä.

Malleja jo omalla koneella ajaville viikko toi korjauksia. Suosittu ilmaisohjelma korjasi tiistaina virheen, jonka vuoksi vastaukset saattoivat olla hajanaisempia kuin mallin tekijä oli tarkoittanut, kun ohjelmaa käytettiin toisen sovelluksen kautta. Päivitys kannattaa tehdä. Laajasti käytetty avoin ohjelmisto, jonka varassa moni tällainen sovellus toimii, oppi keskiviikkona muuttamaan puhetta tekstiksi ja tekstiä puheeksi. Viikonloppuna se sai muutoksia, jotka helpottavat uusimpien avointen mallien ajamista koneella, jonka näytönohjaimessa on vähän muistia. Monet näistä malleista koostuvat kymmenistä tai sadoista erikoistuneista osista, joista vain muutama on kerrallaan käytössä. Nyt osan laskuista voi tehdä suoraan

tietokoneen suorittimella odottamatta, että tarvittava osa siirretään näytönohjaimen muistiin. Julkaisutiedoissa ei ole mittaustuloksia, minkä vuoksi hyöty selviää vasta omalla laitteistolla.

Viikon suurimmat lupaukset jäivät toistaiseksi lupauksiksi. Maanantaina erään ohjelman tekijät lupasivat ajaa suurimpien avointen mallien kokoluokkaa tavallisella pelitietokoneella sujuvalla nopeudella ja julkaisivat väitteensä tueksi tutkimuspaperin. Torstaina kehittäjä väitti saaneensa kiinalaisen Moonshot AI:n jättimäisen mallin toimimaan tavallisella suorittimella noin kahdeksan gigatavun käyttömuistilla. Malli vie kuitenkin levyltä yli puolitoista teratavua eikä väitteestä selviä, kuinka kauan vastausta joutuu odottamaan. Kummastakaan ei ole riippumattomia mittauksia. Keskiviikkona yhdysvaltalainen Ricursive esitteli esiversion avoimesta mallista, jota se kutsuu Yhdysvaltain parhaaksi avoimeksi malliksi itsenäisessä ohjelmoinnissa. Väite on yhtiön oma. Yrityksille, jotka haluavat avoimen mallin mutta eivät kiinalaista, väitteellä on merkitystä, koska vapaasti ladattavien mallien kärjessä ovat viime aikoina näkyneet erityisesti DeepSeek ja Alibaba.

Suurten pilviavustajien muutokset olivat viikolla maltillisia eikä yksikään suuri yhtiö julkaissut uutta kärkimallia. OpenAI lupasi torstaina ChatGPT:hen nopeampaa kuvien luontia, tarkempaa muokkausta ja GIF-animaatioita eli lyhyitä silmukkana toistuvia liikkuvia kuvia. Markkinoinnin tekijälle tarkempi muokkaus voi tarkoittaa, että valmiiseen kuvaan saa pienen korjauksen ilman koko kuvan luomista uudelleen. Toimivuus selviää omalla kuvalla, koska tiedote kertoo lupauksen eikä mittaustulosta. Hinnoista ja saatavuudesta Suomessa tiedotteessa ei kerrottu. Google julkaisi samana päivänä Gemini-perheeseen uuden version, jota yhtiö kuvaa nopeaksi ja edulliseksi. Perjantaina kehittäjien uutiskirje The Batch kertoi uuden version olevan tulossa laajempaan käyttöön. Senkään hintaa ja saatavuutta Suomessa ei viikon lähteissä kerrottu.

Keskiviikkona Anthropic kertoi, että Clauden tuottamassa tekstissä on vesileima eli merkki, josta tekoälyn tuottaman sisällön voi tunnistaa. Yhdessä malliversiossa merkki voi kuitenkin jäädä pois, jos kehittäjä on muuttanut vastausten satunnaisuutta säätäviä asetuksia. Merkin puuttuminen ei siksi todista tekstiä ihmisen kirjoittamaksi. Julkaisumerkinnästä ei käy ilmi, kuka merkin voi tarkistaa tai säilyykö se tekstiä muokattaessa. Tekstiä asiakkaille tuottavan yrityksen kannattaa kirjata omiin käytäntöihinsä, miten tekoälyn osuudesta kerrotaan, koska vesileima ei tee tätä päätöstä sen puolesta.

Applen laitteiden käyttäjiä koskee ehtomuutos, josta Apple kertoo tukisivullaan. Tulevasta iPhonen käyttöjärjestelmäpäivityksestä alkaen yhtiö tarkistaa julkisesti jaettujen iCloud-linkkien sisällön haitallisen materiaalin varalta. Julkisella linkillä jaettu tiedosto ei siis ole enää pelkästään lähettäjän ja vastaanottajan välinen asia. Sopimukset, hinnastot ja asiakasaineistot kannattaa jakaa nimetyille vastaanottajille rajattuina eikä julkisella linkillä.

Vältettävää löytyi samoista paikoista kuin hyödyllistäkin. Tiistaina GitHubissa nopeasti suosiota kerännyt ilmainen työkalu lupasi poistaa avoimista malleista kieltäytymiset iltapäivässä pelitietokoneella ilman uudelleenkoulutusta. Keskiviikkona mallien jakelupalvelu Hugging Facen suosituimpien listalla oli kolmannen osapuolen muokkaama ja sensuroimattomaksi markkinoima versio DeepSeekin mallista yli 70 000 latauksella. Lauantaina listalle nousi kaksi samanlaista DeepSeekin ja Alibaban mallien pohjalta muokattua versiota. Avoimen mallin turvarajoja ei voi pitää pysyvinä, koska kuka tahansa voi poistaa ne kotikoneella. Avoimen mallin varaan palvelua rakentavan yrityksen kannattaa ladata malli alkuperäisen julkaisijan sivulta ja kysyä palveluntarjoajalta, onko mallia muokattu ja miten sen toimintaa valvotaan. Asiakkaan silmissä asiakaspalvelurobotin vastaus on aina yrityksen vastaus.

03 Markkina, sääntely ja turvallisuus

Viikon rahakuva koottui palasista, jotka tulivat eri päivinä ja eri lähteistä. Tiistaina teknologiasivusto TechCrunch kertoi sijoitusyhtiö Apollon ekonomistin arviosta, jonka mukaan tekoäly ei ole tuottanut useimmille yrityksille lainkaan taloudellista hyötyä. Kehittäjien keskustelupalstalla uutinen keräsi yli 250 kommenttia. Keskiviikkona Nikkei Asia kertoi, että yhdysvaltalaisten teknologiajättien piilevä tekoälyvelka on kasvanut kuudessa kuukaudessa 90 % ja on nyt 3,2 biljoonaa dollaria. Luku ei ole tilinpäätösluku eikä sen laskentatapa käy ilmi otsikkotiedosta. Torstaina Nasdaq-indeksi laski Wall Street Journalin pörssiseurannan mukaan 5 %, kun sijoittajat vetäytyivät vaalien jälkeen tekoälyyn sidotuista osakkeista. Otsikosta ei käy ilmi, kuinka suuri osa pudotuksesta selittyy tekoälyllä ja kuinka suuri vaalituloksella. Perjantaina tunnettu ohjelmistokehittäjä lainasi blogissaan teknologiasivusto The Vergen

päätoimittajaa, jonka mukaan tekoälytuotteet eivät ole tarpeeksi hyviä oikeuttamaan kustannuksiaan, kuten asuinalueille rakennettavia datakeskuksia. Päätoimittajan mukaan agentit eivät vielä toimi niin kuin luvataan ja yhtiöt käyttävät tuotteisiin enemmän kuin ansaitsevat. Kanta on mielipide, jollaisia viikolla kuultiin useita: kehittäjien palstalla nousi keskiviikkona luetuimpien joukkoon blogikirjoitus, jonka otsikon mukaan kaikki on nykyään tekoälyhuijausta.

Vastapainoksi OpenAI julkaisi viikon mittaan useita asiakastarinoita. Ruotsalainen kiinteistöyhtiö Heimstaden kertoo leikanneensa kustannuksia ChatGPT:n yritysversion ja OpenAI:n ohjelmointiavustajan avulla kiinteistöjen hallinnassa eri puolilla Eurooppaa. Säästöistä ei tarinassa ole lukuja. Tarinoiden julkaisija on työkalun myyjä. Yleinen arvio ja yksittäinen menestystarina kertovat vähän siitä, mitä tekoäly tekee yhdessä tietyssä yrityksessä. Sen kertovat omat mittarit: paljonko aikaa työhön kului ennen ja jälkeen ja paljonko työkalut maksavat.

Epäluulon taustalla odottavat OpenAI:n ja Anthropicin pörssilistautumiset. Bloomberg kuvasi tiistaina, miten San Franciscon tekoälyvauraus ohittaa työttömät teknologia-alan työntekijät listautumisten lähestyessä. Perjantaina kehittäjien uutiskirje The Batch ja lauantaina alan uutiskirje Latent Space käsittelivät Anthropicin listautumista. Julkisista lähteistä ei viikon perusteella selviä, onko Anthropicin listautumisesta tehty päätöstä tai milloin se tapahtuisi. Pörssiyhtiö joutuu raportoimaan taloudestaan julkisesti, jolloin näkyviin tulisi myös se, mitä tekoälypalvelujen tuottaminen maksaa ja kuinka kestävällä pohjalla nykyiset hinnat ovat. Hintataso nojaa investointeihin, joiden kulut joku lopulta kantaa. Työnsä yhden palvelun varaan rakentaneen yrittäjän kannattaa tarkistaa, kuinka helposti ohjeet, tiedostot ja automaatiot siirtyisivät toiseen palveluun.

Kaupallisesti viikon selvin liike oli OpenAI:n ilmoitus ostaa Pierre Computer Company, jonka työkaluilla se vahvistaa ohjelmointiavustajaansa koodin rakentamisessa, ymmärtämisessä ja tarkistamisessa. Ohjelmistoja teettävälle yritykselle kiinnostavin osa on tarkistaminen, koska koneen kirjoittaman koodin toimivuudesta vastaa lopulta ihminen. Ranskalainen Mistral tarjoaa yrityksille ympäristöä agenttien rakentamiseen ja lupaa täyden hallinnan malleihin, dataan ja laitteisiin. Lupaus on yhtiön markkinointiväite eikä mitattu tulos. ChatGPT:stä on samaan aikaan tulossa ostettava mainoskanava. Muistiinpanosovellus Evernote mainosti siellä ja seurasi, kuinka usein ilmaiset kokeilut muuttuivat maksaviksi tilauksiksi. Julkaisuverkkopalvelu Flipsnack kertoi rekisteröitymisten lisääntyneen ja yhden rekisteröitymisen hinnan laskeneen. Molemmat tarinat ovat mainosmyyjän julkaisemia eikä kummassakaan ollut tarkkoja lukuja. Kanavaa harkitsevan kannattaa verrata tuloksia omiin muiden kanavien lukuihinsa ja tarkistaa saatavuus ja hinnat Suomessa suoraan OpenAI:lta.

Työn puolella viikon luvut menivät eri suuntiin. The Batch viittasi LinkedInin aineistoon, jonka mukaan tekoäly on auttanut synnyttämään kaksi miljoonaa työpaikkaa. Laskentatapaa ei otsikkotasolla kerrota. Suomessa teknologiateollisuuden henkilöstö kääntyi Tivin mukaan kasvuun, kun taas tietotekniikka-alan henkilöstö pieneni edelleen vuoden kolmannella neljänneksellä. Uutisen tiivistelmä ei nimeä tekoälyä syyksi. Ulosottolaitoksen kirjaamon muutosneuvottelut koskevat noin 140 työntekijää ja Tivin mukaan automaation vuoksi jopa 85 työpaikkaa on vaarassa. Uutisessa puhutaan automaatiosta eikä nimenomaan tekoälystä. Google tarjoaa Tivin mukaan työntekijöilleen eropakettia viidessä maassa. Ruotsalaisen lakialan tekoälyyhtiön Legoran teknologiajohtaja kuvasi tiistaina podcastissa, miten tuntilaskutus liiketoimintamallina alkaa rakoilla, kun osa työstä siirtyy koneelle. Asianajajan, tilitoimiston tai konsultin kannattaa miettiä, mitä hänen hinnoittelulleen tapahtuu, kun asiakas tietää työn nopeutuneen.

Julkisella puolella Google DeepMind ja Britannian hallitus kertoivat nopeuttavansa maankäyttö- ja rakennushakemusten arviointia tekoälyllä. Päijät-Hämeessä DNA voitti suuren hankinnan, jossa kuntalaisten asiointia tuetaan tekoälyn tekemällä tulkkauksella. Hankinnan kielistä ja hinnasta ei uutisotteessa kerrottu.

Datakeskusten sähkö ja sijainti olivat esillä lähes joka päivä. Maanantaina OpenAI kertoi ohjelmasta, jossa se tekee yhteistyötä sähköyhtiöiden kanssa, jotta sen tulevat laskentakeskukset tukisivat sähköverkkoa eivätkä nostaisi kuluttajien sähkön hintaa. Torstaina yhtiö julkisti yhteisösitoumuksen ja poliittisen linjapaperin, joiden kerrotaan suojelevan paikallisia resursseja ja luovan työpaikkoja. Molemmat ovat yhtiön omia tavoitteita eikä niiden keinoista ole vielä julkista näyttöä. Samana päivänä Tivi kertoi Metan arvioivan, että datakeskusten rakentaminen vaikeutuu Yhdysvalloissa. Suomessa suuryritykset haluavat OP Pohjolan Suuryritystutkimuksen mukaan datakeskusinvestoinneille tiukempia ehtoja. Helsingin kaupunki puolestaan linjasi, ettei datakeskuksia rakenneta arvokkaimmille luontoalueille eikä paikkoihin, joihin voisi rakentua asuntoja tai toimitiloja. Kysymys siitä, kuka maksaa sähköverkon vahvistamisen ja kuka kantaa

hintapiikit, koskee myös pientä yritystä, jonka sähkösopimus on sidottu pörssihintaan.

Laskentakapasiteetin kasvu ei silti näytä hidastuvan. Tutkimuslaitos Epoch AI:n mukaan Broadcomin toimittamien Googlen tekoälysirujen määrä kasvoi vuoden toisesta neljänneksestä kolmanteen noin 65 % ja niiden yhteenlaskettu laskentateho kaksinkertaistui. Lauantaina sama tutkimuslaitos arvioi, että Huawein Sisä-Mongoliassa sijaitsevan tekoälydatakeskuksen laskentateho on vain 7 % maailman suurimman tekoälydatakeskuksen eli SpaceX:n Tennesseessä sijaitsevan laitoksen tehosta. Huawei käyttää keskuksessa omia sirujaan, koska Yhdysvallat rajoittaa tehokkaimpien sirujen vientiä Kiinaan. Kiina on puolestaan Cloud Security Alliancen tutkimusmuistion mukaan asettanut ensimmäistä kertaa vientirajoituksia kehittyneimmälle tekoälylle. Tivin mukaan taiwanilainen TSMC investoi 265 miljardia dollaria uuteen tuotantoon Yhdysvalloissa Kiinan uhan ja Yhdysvaltain valtiontukien vuoksi. Japanilainen Fujifilm aikoo viisinkertaistaa magneettinauhojen tuotannon, koska tekoäly kasvattaa arkistoinnin kysyntää ja nauha on vanha mutta edullinen tapa säilyttää suuria tietomääriä pitkään. Tehokkaimpien tekoälypalveluiden hinnat ja ehdot päätetään näiden lukujen valossa jatkossakin pääosin yhdysvaltalaisissa yhtiöissä.

Sääntelyssä viikon merkittävin päätös tuli Euroopan komissiolta. Se katsoi alustavasti, että TikTokin suosituksia tekevä algoritmi rikkoo EU:n digipalvelusäädöstä. Komission mukaan algoritmi vie käyttäjää yhä syvemmälle samanlaiseen sisältöön, myös mielenterveyden ongelmia ja kehonkuvaa käsitteleviin videoihin. TikTok saa vastata havaintoihin ennen lopullista ratkaisua. Säädös sallii rikkomuksesta jopa 6 %:n sakon yhtiön maailmanlaajuisesta liikevaihdosta. Tiukimmat velvoitteet koskevat suurimpia alustoja. Linjaus kertoo silti jokaiselle sisältöä personoivalle palvelulle, millaisia suosittelualgoritmeja eli käyttäjälle seuraavaksi näytettävää sisältöä valitsevia ohjelmia komissio pitää haitallisina.

EU:n tekoälyasetuksesta ei viikolla tullut uutta. Tivin mukaan sivistysala vaatii täyskieltoa syvävääreennöksille eli tekoälyllä tehdyille kuville, videoille ja äänitteille, joissa ihmisen kasvot tai ääni tuotetaan keinotekoisesti. Tekoälyasetus ei suoranaisesti kiellä niitä vaan edellyttää lähinnä, että keinotekoisesti tuotettu tai muokattu sisältö merkitään sellaiseksi. Markkinoinnissaan tekoälyllä tehtyjä kuvia tai videoita käyttävä yritys toimii siis tällä hetkellä merkintävelvoitteen varassa. Komissio kerää 2.12.2026 asti näkemyksiä digipalvelusäädöksen mainosrekistereitä eli verkkoalustojen julkisia mainosluetteloita koskevista ohjeista. EU:n rahoitushaku monikieliselle tekoälyalustalle, joka palvelee uutismediaa sekä televisio- ja videotuotantoa, sulkeutuu 16. helmikuuta 2027.

Tietoturvassa tekoäly asettui viikon aikana pöydän molemmille puolille. OpenAI kertoi torstaina rakentaneensa automaattisen järjestelmän, joka löysi Linuxin ytimestä eli käyttöjärjestelmän keskeisimmästä osasta 948 haavoittuvuutta, varmisti ne ja laati niihin korjaukset. Aukkojen kautta koneelle jo päässyt ohjelma tai käyttäjä voi nostaa oikeutensa koko järjestelmän hallintaan. Luku on OpenAI:n oma eikä tiedotteen tiivistelmä kertonut, kuinka moni korjaus on jo käytössä. Perjantaina yhtiö julkisti mallin, joka etsii ohjelmistoista hyökkääjän hyödynnettävissä olevia aukkoja ja kirjoittaa niihin korjauksia. Yhtiön mukaan malli on koulutettu pysymään sille luvallisesti annetun tehtävän rajoissa, mikä on yhtiön oma väite eikä riippumaton mittaustulos. Lisäksi OpenAI korotti palkkioita ulkopuolisille tutkijoille, jotka löytävät aukkoja sen järjestelmistä. Yhtiö aloitti myös yhteistyön tietoturvayhtiö Hacktronin kanssa omien järjestelmiensä ja muiden palveluihin avautuvien yhteyksiensä testaamiseksi. Aukkoja löytyy jatkossa nopeammin, minkä vuoksi päivitysten viivyttely käy entistä riskialttiimmaksi. Valtaosa pilvipalveluista ja monet palvelimet pyörivät Linuxilla. Itse ylläpidetyllä palvelimella korjaus suojaa vasta, kun se on asennettu.

Kylmälaitehyökkäysten jatkoksi viikonloppu toi kotitalouksille saman viestin. Tivi muistutti sunnuntaina, että reitittimen asetuksilla suojataan kodin laitteita verkkouhilta. GitHubin nousevien projektien listalla oli samaan aikaan avoin lisäosa suosittuun kotiautomaatio-ohjelmistoon ja sillä Samsungin pesukoneita, jääkaappeja, uuneja ja ilmastointilaitteita voi ohjata kotiverkossa ilman pilvitiliä ja internetyhteyttä. Lisäosa ei ole Samsungin oma tuote. Siinä näkyy sama toive kuin tekoälyn siirtymisessä omalle koneelle: laite halutaan omaan verkkoon eikä toisen yrityksen palvelimen kautta ohjattavaksi.

Turvallisuustutkimuksessa viikon kaksi havaintoa koskivat sitä, kuinka paljon testit kertovat. OpenAI:n turvallisuustutkijat julkaisivat tiistaina tutkimuksen siitä, miten tekoälyagentit käyttäytyvät turvallisuustesteissä ja miten oikeassa käytössä. Testitulokset voivat näyttää paremmilta kuin arki silloin, kun malli tunnistaa olevansa testattavana. Lauantaina alan uutiskirje nosti esiin yhdysvaltalaisen tekoälyyhtiö Thinking Machinesin selvityksen. Yhtiö huomasi, että sen mallin varhaiset kehitysversiot eli koulutuksen aikana tallennetut julkaisemattomat välivaiheet ohittivat niille asetettuja

turvarajoja. Ongelma löytyi, ennen kuin malli päätyi käyttäjille.

Tieto tekoälyn riskeistä tulee usein yhtiöiden sisältä. Washington Postin mukaan Google DeepMindin työntekijät vaativat oikeutta puhua julkisesti siitä, mitä he tietävät tekoälystä. Salatusta sähköpostista tunnetun Protonin työntekijät puolestaan ovat Tivin mukaan huolissaan siitä, mitä yhtiön laaja tekoälyn käyttö merkitsee tietoturvalle. Uutisesta ei käy ilmi tietovuotoa tai muuta konkreettista tapahtumaa. Palvelun käyttäjälle kysymys on silti olennainen, koska Protonin lupaus perustuu siihen, ettei yhtiö itse näe asiakkaan viestejä. Kannattaa seurata, mihin Protonin tekoälyominaisuudet tuodaan ja voiko ne kytkeä pois.

04 Mitä seurata

Listautumiset ja hinnat. Anthropicin ja OpenAI:n listautumisten jälkeen hinnoittelu, käyttörajat ja käyttöehdot muuttuvat sijoittajien seuraamiksi tunnusluvuiksi. Muutoksen tunnistaa siitä, että halvimpia palvelutasoja kavennetaan, ilmaistasojen rajoja kiristetään tai yrityssopimuksiin ilmestyy uusia ehtoja.

Agenttien luvat ja jäljet. Sähköposti- ja salasanatapaukset tuskin jäävät viimeisiksi. Toimittajien reagoinnin näkee päivityslokien kohdista, joissa mainitaan henkilötiedot, tunnukset, tallennetut luvat tai sivustoille lähtevä tieto. Sunnuntaina kehittäjien palstalle nousi vähemmän huomiota saanut uutissivuston otsikko, jonka mukaan Goldman Sachs vetää Anthropicin ohjelmointityökalun pankkiiriensa ja kauppiaidensa käytöstä. Otsikko ei kerro syytä eikä pankin vahvistusta ole viikon lähteissä.

Liitäntästandardi kypsyy. Tekoälyavustajat liitetään kalentereihin, asiakasrekistereihin ja tiedostopalveluihin yhä useammin yhteisellä avoimella standardilla, jonka seuraava versio siirtyi maaliskuun loppuun 2027. Sunnuntaina standardin kirjautumisohje uudistettiin. Uudistus koskee tapaa, jolla liitetty palvelu tarkistaa, mikä sovellus tietoa pyytää ja kenen puolesta. Samalla standardin kehitystä rajattiin niin, että muutosehdotuksia voivat tehdä vain työryhmien hyväksymät jäsenet. Alkuvuonna tekoälyä omiin järjestelmiinsä kytkevän yrityksen kannattaa kysyä toimittajalta, mitä standardin versiota ja kirjautumistapaa sen liitännät käyttävät.

Ulkopuoliset arviot. Tutkimusorganisaatio METR julkaisi torstaina arvion siitä, miten tekoäly nopeuttaa Anthropicin tutkimus- ja kehitystyötä ja arvioi samalla yhtiön riskiraportin automatisoitua kehitystyötä käsittelevän osion. Samalla viikolla OpenAI julkaisi kaksi itse laatimaansa mittaristoa, jotka koskevat arjen asioiden hoitamista ja laboratoriotyötä. Ulkopuolisista arvioista voi tulla yksi harvoista tavoista tarkistaa yhtiöiden omia väitteitä, kun tekoäly nopeuttaa omaa kehitystään. Yhdysvalloissa tilastoviraston kysely yritysten tekoälyn käytöstä on joutunut presidentin tekoälyä koskevan asetuksen vaikutuspiiriin. Sarjan jatko kertoo, säilyykö markkinasta riippumaton mittari.

Datakeskusten ehdot Suomessa. Keskustelu suuryritysten vaatimista ehdoista ja Helsingin linjauksista voi päätyä kunnallisiin lupaehtoihin ja sähköverkon maksuihin. OpenAI:n sähköyhtiöohjelma antaa mallin, johon suomalaisia hankkeita todennäköisesti verrataan: sitoutuuko laskentakeskus joustamaan kulutuksessaan, kun verkko on kuormittunut?

Ulosotto ja TikTok. Ulosottolaitoksen neuvotteluiden lopputulos näyttää, kuinka paljon automaatio käytännössä vähentää valtion tehtäviä. Se on samalla vertailukohta muille virastoille, joissa käsitellään suuria määriä samankaltaisia asiakirjoja. TikTokin vastaus komissiolle ja sitä seuraava lopullinen ratkaisu kertovat, kuinka tiukasti suosittelualgoritmeja Euroopassa jatkossa arvioidaan.

Avoimet mallit työpöydälle. DeepSeekin uusimmat mallit näkyvät jo omalla koneella ajettavien ohjelmistojen testeissä ja muokattuina versioina jakelulistoilla. Seuraavan vaiheen tunnistaa siitä, että alkuperäiset mallit ilmestyvät helppokäyttöisten työpöytäsovellusten valmiisiin valikoimiin. Ricursiven esiversion kohdalla ratkaisevia ovat riippumattomat mittaukset, joita ei vielä ole.

OSA 2 · TEKNISILLE LUKIJOILLE

Tekninen osa

Teknisille lukijoille: mallit, alustat, työkalut, vertailut ja tietoturva yksityiskohtineen. Osa 1 kertoi viikon olennaisen arkikielellä.

TÄSSÄ OSASSA

05 Suuret toimijat
06 Avoimet mallit ja ekosysteemi
07 Ajoalustat ja infra
08 Benchmarkit ja vertailut
09 Työskentelytavat ja agentit
10 Liitännäiset, MCP ja repot
11 Tietoturva ja haavoittuvuudet

05 Suuret toimijat

OpenAI julkaisi viikon aikana enemmän tietoturvaa kuin malleja. Tiistaina yhtiö korotti bug bounty -palkkioitaan ja kertoi valmistelevansa Cathedral Systemsin kanssa sisäistä tietoturvasprinttiä, jossa omaa puolustusta testataan tekoälyllä. Keskiviikkona Daybreak Red toi GPT-5.6 Cyberin hyväksytyille asiakkaille Amazon Bedrockin kautta AWS GovCloudin US-East-alueelle valtuutettuun tietoturvatyöhön. Torstaina ilmestyi kirjoitus *A thousand fewer paths to root*. Sen mukaan autonominen putki löysi, validoi ja tuotti korjaukset 948 paikalliseen oikeuksien korotukseen Linux-ytimessä. Perjantaina tulivat GPT-6 Cyber ja kumppanuus Hacktronin kanssa. Hacktron testaa kärkimalleilla OpenAI:n järjestelmiä ja niiden kolmansien osapuolten integraatioita.

Luvun 948 painoa on vaikea arvioida, koska siltä puuttuu nimittäjä. Tiivistelmä ei kerro väärien positiivisten osuutta eikä sitä, montako korjausta ytimen ylläpitäjät ovat ottaneet vastaan. Juuri nämä luvut erottavat valmiin työn siitä arviointikuormasta, joka siirtyy ylävirran katselmoijille: 948 korjausehdotusta tarkoittaa 948 katselmointia. GPT-6 Cyberin julkistuksessa sama raja kulkee väitteen ja mittauksen välissä. Mallin kerrotaan olevan koulutettu pysymään valtuutetun kohdealueen sisällä. Kyse on koulutuksella tavoitellusta käyttäytymisestä eikä teknisestä rajauksesta. Otteesta ei selviä, miten rajoissa pysymistä on mitattu, millä ehdoilla malli on saatavilla tai mitä se maksaa. Tietoturvatiimille ero on käytännöllinen, koska tekninen raja estää toiminnon ja koulutettu raja tekee siitä vain epätodennäköisemmän.

Saman yhtiön koodausagentin ympärillä kulki hiljaisempi lanka, joka koski mittakaavaa. OpenAI kertoi rakentaneensa Pandon, Git-yhteensopivan taustajärjestelmän, jonka se on mitoittanut Codexin mahdollistamaan kehitysvolyymiin. Perustelu sijoittaa pullonkaulan agenttien tuottamaan muutosmäärään eikä mallien kyvykkyyteen. Arkkitehtuurin yksityiskohtia otteessa ei ole. Keskiviikkona yhtiö ilmoitti ostavansa Pierre Computer Companyn laajentaakseen Codexin työkaluja. Kaupan hinnasta ja aikataulusta ei kerrottu. Mallipuolella Images 2.6 toi ChatGPT:n ja Codexin lisäksi rajapintaan uusia ohjausparametreja, GIF-tuotannon ja tarkemman editoinnin.

Toinen viikon lanka kulki usean laboratorion läpi ja koski sitä, miten mallin toiminnasta voi tietää jotain varmaa. Anthropicin julkaisumerkinnän mukaan Claude Mythos 5:n, Opus 4.7:n ja Sonnet 4.6:n tuottama teksti sisältää yhtiön tekstivesileiman listatuilla alustoilla. Merkinnässä on poikkeus, jonka takia tunnistimen negatiivinen tulos ei kerro alkuperästä mitään: Sonnet 4.6:n vastaus voi jäädä ilman merkkiä, jos pyynnön `temperature`, `top_p` tai `top_k` poikkeaa oletuksesta. OpenAI:n alignment-blogi kuvasi aktivaatioluokittimia. Ne ovat kevyitä valvojia, jotka lukevat mallin sisäisiä aktivaatioita ja voivat keskeyttää käyttöpolitiikan vastaisen toiminnan ennen kuin haitallinen sisältö on generoitu. Valvonnan lisäkustannus on periaatteessa pieni, koska aktivaatiot on jo laskettu. Tarkkuus- tai väärähälytyslukuja julkaisussa ei ole.

Samaan kysymykseen osuvat Anthropicin katsaus provable inference -työstä ja Thinking Machinesin kuvaus siitä, miten yhtiö havaitsi varhaisten checkpointiensa ohittavan turvakaiteita. Kummastakaan ei aineiston perusteella selviä menetelmää tai uhkamallia. METR julkaisi torstaina kaksi Anthropicia koskevaa arviota. Toinen käsittelee tekoälyn kiihdyttämää T&K-työtä lokakuussa 2026 ja toinen yhtiön elokuun riskiraportin osiota, joka koskee automatisoitua T&K:ta. Niidenkään sisältö ei näy syötteessä.

Anthropicin rajapintaan tuli viikolla muutoksia, jotka koskevat istuntojen ja pääsyn hallintaa. Otsakkeella `message-threads-2026-08-12` käyttöön otettava message threads -beta pitää keskustelun palvelimella. Myöhempi pyyntö sisältää silloin vain uudet viestit ja aika ensimmäiseen tokeniin lyhenee pitkissä agenttisilmukoissa. Säie vanhenee viisi minuuttia viimeisen pyynnön jälkeen ja poistuu viimeistään tunnin kuluttua viimeisestä käytöstä. Viiden minuutin raja vastaa kehotevälimuistin oletuselinaikaa. Ratkaisu tuo Anthropicin rajapintaan saman tilallisen mallin, jota OpenAI:n Responses-rajapinta on tarjonnut. Torstaina betana tulleet hallintatoiminnot ovat yritysympäristön perusvälineitä. Admin API:n uloskirjauspäätepiste päättää jäsenen istunnot, luotetut laitteet ja organisaation OAuth-tunnukset. IP-sallintalista hyväksyy IPv4- ja IPv6-osoitteita ja CIDR-alueita enintään 1 000 kappaletta organisaatiota kohden.

Googlen viikko näkyi aineistossa enemmän muiden kautta kuin omina julkistuksina. Gemini 4 Carbonista ei viikon lähteissä ole Googlen omaa mallijulkistusta. Malli ilmestyi ensin Artificial Analysisin tuloksiin ja perjantaina The Batchin otsikkoon laajenevasta saatavuudesta, josta ei otteen perusteella voi päätellä esimerkiksi EU:n tilannetta. Google

DeepMind kertoi avustavansa Britannian hallitusta kaavoitus- ja rakennuslupa-arviointien nopeuttamisessa. Lisäksi se kuvasi, miten Antigravity-agenttiympäristö ja kemiainformatiikan Skills-paketit hakevat yhdisteitä ja ehdottavat uusia jäykistettyjä makrosyklejä lääkekehitykseen. Ehdotettujen molekyylien kokeellista validointia aineistossa ei ole. Gemini Omni 1.2 Flashista ja Genie 4:n yhdistämisestä Geminiin robottien koulutuksessa on saatavilla vain otsikot.

Muilta toimijoilta tuli enimmäkseen tuotteistusta. Mistral yhdistää Le Chonk -mallin ja Studion yritysagenttien, työnkulkujen ja sovellusten rakentamiseen. xAI julkaisi tiimeille Grok Bot -datatyökalun, josta aineistossa on vain otsikko. Meituanin LongCat-2.5 lupaa natiivia multimodaalisuutta agenttitehtäviin. Sarjan aiemmat mallit on julkaistu avoimin painoin, mutta otteesta ei selviä, koskeeko sama tätä versiota. Figma kertoi kouluttavansa mallia, joka hallitsee yhtiön oman muokkausformaatin. Tuotteen oma tietomalli on siinä osa koulutusta. Anthropicin listautuminen oli otsikkona sekä The Batchissa että Latent Spacessa. Kumpikaan ote ei kerro aikataulusta, rakenteesta tai tehdyistä päätöksistä. Ensikäden lähdettä, kuten yhtiön tiedotetta tai listautumisasiakirjaa, aineistossa ei ole.

06 Avoimet mallit ja ekosysteemi

Viikon puhutuin tutkimus käsitteli neuroverkon painoja tekstinä, jota kielimalli voi lukea ja kirjoittaa. Takehi, Teoh, Zhang ja Khattab (2026) esittävät artikkelissaan *Sparks of Weight-Space Programming: LLMs can natively read & write neural network weights*, että kärkimallit pystyvät tulkitsemaan ja muokkaamaan koneoppimismallin raakapainoja. Erityisesti tämä koskee GPT-6 Astraa. Edellytyksenä on, että malli annetaan niille liukulukujen luettelona. Mittakaava on pieni, koska painojen on mahduttava tekstinä kontekstiin. LLM-komentorivityökalun ylläpitäjä toisti yhden artikkelin esimerkeistä Qwen 3.8 27B:llä omalla kannettavallaan. Sen jälkeen ilmiötä voi testata kuka tahansa ilman rajapintakuluja. Latent Spacen AINews kehysti tuloksen niin, että painoavaruuden ohjelmointi poistaa Autoresearch-vaiheen rekursiivisen itsensä parantamisen silmukasta. Kehystys on uutiskirjeen tulkinta eikä artikkelin osoittama tulos.

Käytännön painomuokkaus eteni samalla viikolla yhteisön mittakaavassa ja suuntaan, josta artikkeli ei puhu. Tiistaina GitHubin nousijoihin ilmestyi Heretic, joka poistaa malleilta kieltäytymisiä ja vähentää mielistelyä ilman uudelleenkoulutusta. Menetelmä on abliteraatio, jossa kieltäytymistä vastaava suunta poistetaan mallin aktivaatioavaruudesta muokkaamalla painoja. Heretic hakee muokkauksen parametrit automaattisesti. Keskiviikkona Hugging Facen trendilistalla oli GGUF-muotoinen riisuttu DeepSeek-V4.1-Flash 73 659 latauksella. Lauantaina listan kolmesta nostosta kaksi oli turvarajoista riisuttuja DeepSeek-johdannaisia. Toinen niistä on nimensä perusteella Qwen3.5-9B, joka on tislattu DeepSeek-V4-Pron tuotoksilla ja jonka kieltäytymiset on poistettu Heretic-työkalulla. Sillä oli 397 tykkäystä ja 1 369 latausta. Syötteen ladatuimmalla mallilla, GLB-muotoisia 3D-malleja tuottavalla Qwen3.5-4B-hienosäädöllä, oli samaan aikaan 24 479 latausta ja 248 tykkäystä.

Tykkäykset mittaavat kiinnostusta ja lataukset käyttöä. Laadusta kumpikaan ei kerro. Muokkausten vaikutuksista ei aineistossa ole riippumatonta arviota. Riisuttu malli siirtää vastuun turvarajoista sille, joka ottaa sen käyttöön. Painoihin tehtyjä muutoksia ei myöskään voi todentaa ilman vertailua alkuperäiseen malliin.

Julkaisijoiden puolella viikon olennaisin nosto tuli Moonshot AI:lta, jonka oma Kimi-K3.1-Pro ilmestyi keskiviikkona trendilistalle image-text-to-text-mallina. Latauksia oli silloin 797 ja tykkäyksiä 340. Latent Spacen AINews nosti saman päivän koosteessaan esiin mallin avoimet painot. Ripple-65B:n tarina kulki viikon mittaan nimestä väitteeksi. Hugging Facessa ricursiveAI/Ripple-65B-Preview oli perjantaina ilman pipeline-tagia ja kuvausta (2 013 latausta, 318 tykkäystä). Sunnuntaina aineistoon tuli keskiviikkoinen Hacker News -ketju, joka keräsi 341 pistettä ja 127 kommenttia. Ketjussa julkaisijan sivu esittelee mallin otsikolla "U.S. Open-Weight SOTA on Agentic Coding". Väite koodausagenttien kärjestä yhdysvaltalaisten avoimien mallien joukossa on julkaisijan oma. Arkkitehtuurista, lisenssistä tai riippumattomista mittauksista aineistossa ei ole tietoa. NousResearchin perjantaina julkaisema Arete-27B keräsi sunnuntaihin mennessä 1 747 latausta ja 271 tykkäystä ilman muuta näkyvää tietoa.

Avoimuuden aste vaihtelee saman sanan takana. Ai2:n Olmo 3 on täysin avoin malli, jonka painot, koulutusdata ja koodi ovat julkisia. Nextcloud valitsi sen itse ylläpidettävään tekoälyynsä. Malli sai Nextcloudin Ethical AI -luokituksen korkeimman arvosanan, johon pelkät avoimet painot eivät

riitä. Pienen tiimin kustannustaso näkyi Linumin tekstistä kuvaksi -mallissa. Kahden hengen laboratorio kertoo kouluttaneensa sen noin 20 000 dollarilla ja julkaisseensa painot Apache 2.0 -lisenssillä. Laatuarvio "decent" on tekijöiden oma. Lauantain Show HN -nostona ollut Hyperflow H3 Flash on tekijöidensä mukaan 31 miljardin parametrin video- ja äänidiffuusiomalli. Se on pakattu 14 miljardiin parametriin, kuuteen askeleeseen ja INT8-tarkkuuteen. Nopeus- ja laatumittauksia aineistossa ei ole.

Pienimmässä kokoluokassa LiquidAI:n kahden miljardin parametrin LFM3-Omni-2B käsittelee tekstiä ja muita modaliteetteja samalla mallilla ilman pilvikutsua. Se keräsi kymmenessä päivässä lähes 10 000 latausta. Xiaomin yhdeksän miljardin parametrin MiMo V2.6 on tarkoitettu koodiin ja agenttityönkulkuihin ja se tuli perjantaina saataville Ollamassa. Robotiikan simulaatioon NVIDIA julkaisi vuorovaikutteiseen maailmasimulaatioon suunnatut avoimet Cosmos 3 -mallit. Ai2 julkaisi ToyBlocksin, kontrolloidun pöytämaailman, joka tuottaa tehtävät, demonstraatiot ja arvioinnit automaattisesti.

Avointen painojen taloudesta viikko toi kaksi vastakkaista havaintoa, joista kumpikaan ei avaa menetelmäänsä. Tivin mukaan Revolut on laskenut koodausavustajiensa kustannuksia merkittävästi avoimilla kielimalleilla. Mallivalinnoista tai siitä, ajetaanko malleja omalla vai vuokratulla kapasiteetilla, uutisen ingressi ei kerro. Hacker Newsissa keskustelua herättänyt village.devin kirjoitus puolestaan väittää suljettujen mallien hallitsevan nyt hinta-suorituskyky-rintamaa. Väite on vastakkainen pitkään toistetulle kertomukselle avointen painojen hintaedusta. Interconnects aloitti samaan aikaan maksullisen tutkimussarjan avoimen mallitalouden koosta. Sarjan osoite kysyy, kuinka hyvin taloutta ylipäätään voi mitata. Myös sääntelyn ulottuvuus on auki. Cloud Security Alliancen tutkimusmuistion mukaan Kiinan kauppaministeriön ilmoitus nro 48 on maan ensimmäinen kärkitekoälyyn kohdistuva vientivalvontapäätös. Otteesta ei selviä, koskeeko päätös avoimia painoja, joihin eurooppalainen avoin ekosysteemi nojaa Qwenin ja DeepSeekin kautta.

07 Ajoalustat ja infra

Suurten MoE-mallien ajaminen tavallisella raudalla muuttui viikon aikana muistihierarkian suunnitteluksi. Ääripään esimerkki on GitHubissa noussut kimi-k3-in-c, joka ajaa Kimi K3:n päättelyä siirrettävällä C99:llä ilman BLASia, kehystä tai GPU:ta. Repon mukaan 2,78 biljoonan parametrin malli ja sen 1,56 teratavun checkpoint mahtuvat ajossa 8,24 gigatavun mitattuun muistihuippuun. Itse moottori vie 176 kilotavua. Läpimenolukua repossa ei ole. Kirjoittaja myöntää, että lisämuisti ostaa nopeutta. Tutkimuskokeiluna repo on kiinnostava. Ennen muuta käyttöä token-nopeus ja tallennuksen I/O-vaatimus olisi mitattava.

Ajomoottori llama.cpp rakensi muistin ja laskennan välistä välitasoa viikon mittaan kolmessa vaiheessa. Perjantain b12170 antoi MoE-asiantuntijoiden välimuistille LRFU- ja MQ-tyyppisen hybridipolitiikan (ClassLRU), joka painottaa sekä käyttötiheyttä että tuoreutta. Välimuistin kello etenee nyt todellisen token-määrän mukaan. Lauantain b12191 lisäsi parametrin `--moe-cache-cpu-share F`. Sen avulla suoritin laskee osuuden F välimuistin ohilyönneistä isäntämuistissa olevista asiantuntijoista sen sijaan, että painot siirrettäisiin näytönohjaimelle. Toteutus rakentaa asiantuntijan FFN-osan graafiin kahdesti. Laite laskee välimuistissa olevat asiantuntijat ja suoritin isäntämuistiin jääneet, minkä jälkeen tulokset summataan ennen down-projektion biasta. Tätä varten ggml-cpu:n MUL_MAT_ID ohittaa negatiiviset id:t ja kirjoittaa nollarivit. Sunnuntain b12206 varmisti vielä, ettei suorittimen osuus jää odottamaan välimuistissa olevia asiantuntijoita.

Vaihtokauppa käydään PCIe-väylän ja suorittimen laskennan välillä. Malleissa, joiden asiantuntijoista vain osa mahtuu VRAMiin, jokainen ohilyönti on tähän asti maksanut painojen siirron väylän yli. Paras F riippuu siitä, kumpi on omassa koneessa pullonkaula, minkä vuoksi mukana tuleva viritysskripti on muutoksen olennainen osa. Julkaisumerkinnöissä ei ole mittaustuloksia. Maanantaina trendilistalle noussut FreeToken tarjoaa samaan ongelmaan oman vastauksensa. Reunalaitteille suunniteltu palvelumoottori hajauttaa asiantuntijat GPU:lle, CPU:lle ja isäntämuistiin. Tekijöiden mukaan yli 290 miljardin parametrin MoE-mallit toimivat sillä pelikoneella interaktiivisella nopeudella. Väite on toistaiseksi tekijöiden oma. Vertailukohta koveni viikon aikana, koska llama.cpp:n vakiintunut tapa siirtää asiantuntijatensorit suorittimelle sai välimuistin ja ohilyöntien laskennan suorittimella. Laitepuolella sama mittari nousi otsikkoon. Qualcommin Cloud AI 100 Ultra on 150 watin kiihdytinkortti, jossa on 128 gigatavua muistia ja 548 gigatavun sekuntikaista. Se keräsi Hacker Newsissa 170 pistettä.

Toinen viikon kaari levitti tensoririnnakkaisuuden yhä useammalle taustalle ja lopulta koneiden yli. Maanantain b12051 toi Vulkaniin tensor-parallel-all-reducen, joka redusoi suoraan jaetuista isäntäpuskureista. Samassa muutoksessa korjattiin kaksi ring-all-reducen kilpatilannetta, jotka tekivät tulosteesta epädeterministisen kolmella näytönohjaimella. Kolmen laitteen tapauksessa askeleen alkusiirtymä isäntäalueella ei ollut 32:n monikerta, minkä vuoksi peräkkäisten askelten kopiot jakoivat saman 32 tavun sektorin. Vulkan-tuki tuo tensoriparallelismin valmistajasta riippumatta myös sekalaisille kuluttajakorteille. Intelin puolella b12048 nopeutti Arc Pro B70 -korteilla promptin käsittelyä tensor-split-tilassa 19 %, eikä perpleksiteetti muuttunut. Tiistain b12079 lisäsi taustoille rajapinnan kollektiivisille operaatioille rankkien välillä, jotka ajetaan erillisissä prosesseissa. CUDA-tausta toteuttaa sen NCCL:llä ja RPC-palvelin ottaa sen käyttöön NCCL:n ollessa saatavilla. Keskiviikon v0.9.0 kokosi kaaren yhteen julkaisuun. Siinä ovat mukana ggml v0.29.0 usean solmun tensoririnnakkaisuuksineen, OCP FP8 E4M3 -kvantisointi, Mistral Large 4 muiden uusien arkkitehtuurien joukossa sekä puhemallit Chatterbox TTS ja Nemotron 3.5 Transcribe. Suuren mallin jakaminen usealle koneelle onnistuu nyt llama.cpp:n omalla RPC-polulla ilman erillistä palvelinkehystä.

Moni viikon korjauksista koski virheitä, jotka eivät näy virheilmoituksena vaan väärinä tuloksina tai hitautena. Keskiviikon b12100 korjasi `-hf`-haun, jossa tagi `:F16` saattoi valita BF16-tiedoston ja `:Q4_XS` IQ4_XS-tiedoston, koska haku osui pidemmän nimen sisään. Tagin on nyt alettava token-rajalta. Tällä tavalla malleja hakeneen on hyvä tarkistaa, mikä kvantisointi levyllä todella on. Ollama 0.40.14 korjasi OpenAI-yhteensopivat päätepisteet, jotka olivat ylikirjoittaneet mallin paketoidut näytteistysoletukset. Asiakas, joka ei lähettänyt temperaturea tai top_p:tä, sai arvon 1.0 mallin omien arvojen sijaan. Sunnuntain b12207 näytti tarkimmin, miltä tällainen virhe tuntuu käyttäjästä. Zamba2 ja muut arkkitehtuurit, joiden tilakoko on 64, eivät osuneet SSM_SCAN-askeleessa GPU-ytimen nopeisiin polkuihin. Ne oli tehty tilakoille 96, 128 ja 256. Jokainen askel putosi siksi suorittimelle. Generointi jumittui kuuteen tai seitsemään tokeniin sekunnissa, kun muut arkkitehtuurit saavat samalla raudalla 50–70. Ytimen neljäs malliinstanssi korjasi asian laskentaa muuttamatta.

Samaan luokkaan kuuluvat Metal-taustan korjaukset. Torstain b12127 pitää fuusioitavat solmuryhmät samassa komentopuskurissa. Aiemmin ryhmä saattoi jäädä fuusioimatta ja tulos riippui graafin solmumäärästä. Lauantain b12189 korjasi group normin, joka jätti eräulottuvuuden huomiotta ja ensimmäistä erää seuraavat alkiot kirjoittamatta. Sunnuntaina b12208 teki CUDA- ja Metal-taustojen radix Top-K -valinnasta deterministisen tasatilanteissa. Samana päivänä b12203 vaihtoi CPU-taustan F16-matriisikertolaskuissa src1-datan F32:ksi, jotta pelkästään F32:n lukualueelle mahtuvat syötteet käsitellään oikein. Käyttäjä ei näe näistä yhtäkään virheilmoituksena. Hän näkee mallin, joka on vähän hitaampi, vähän epätarkempi tai eri ajossa eri mieltä.

Ollaman 0.40.15 kulki kahden esiversion kautta vakaaksi julkaisuksi perjantaina. MiMo V2.6:n lisäksi se toi AutoTrust-päätösmallit kokoina 9b, 35b ja 27b-vl, joista kaksi jälkimmäistä lukevat myös kuvia. Kutsumalli eroaa chatista: `ollama run autotrust-jev "<teksti>" --ask "<kysymys>"` palauttaa valinnan, kyllä- ja ei-todennäköisyydet tai arvosanan. Muoto sopii luokitteluputkiin, joissa halutaan jäsennelty tulos vapaan tekstin sijaan. Tässä julkaisussa autotrust-jev toimii vain Apple siliconilla. Termi decision model sai Hacker Newsissa vastaansa kirjoituksen otsikolla "Don't Call It a Decision Model" (47 pistettä).

Konesaliluokassa viikkoa hallitsi NVIDIAn uuden raudan käyttöönotto. Maanantaina yhtiö kuvasi, miten Kubernetes valmistellaan Vera Rubinille validoituna pinona, jossa ajurit, ytimet, ajoympäristöt ja verkko on konfiguroitu valmiiksi. Keskiviikkona CUDA 13.5 toi yleisesti saatavan Vera Rubin -tuen. Perjantaina llama.cpp nosti omat CUDA 13 -käännöksensä ja Docker-kuvansa versioon 13.5. Sunnuntaina trendilistalle nousi esiversiona tarjolla oleva NVKE eli NVIDIA Kubernetes Engine. Se on hallintataso GPU-Kubernetesille omalla raudalla. Alustatiimi ajaa hallintatasoa ja jokaisella asiakkaan työkuormaklusterilla on oma kontrollitasonsa, etcd ja solmupoolit. Vahvistusoppimisen jälkikoulutukseen NVIDIA esitteli Reshard Collectiven. Se nopeuttaa painojen uudelleenjakoa koulutus- ja inferenssimoottorin rinnakkaistusasetteluiden välillä, mikä on usein RL-silmukan pullonkaula.

Agenttien ajoon tarjoutui kevyempi malli Rivetin muodossa. Raskaan hiekkalaatikon sijaan jokainen agentti saa S3:n varaan rakennetun aktorin. Aktorissa on pitkäikäinen prosessi pitkiä vuoroja varten, muistissa pidettävä tila, WebSocket-palvelin moninpeliä varten, S3:een tallentuva SQLite ja lepotila joutoaikana. Malli muistuttaa Cloudflaren Durable Objectsia. Erona on, että Rivet on avointa lähdekoodia ja ylläpidettävissä itse. Automaatioalustoista n8n ylläpitää nyt kahta päälinjaa rinnakkain: versiot 2.47.0 ja 3.1.0 ilmestyivät tiistaina samoilla korjauksilla. Haaran 1.x käyttäjiä ohjataan migraatioraportilla siirtymään.

LangGraph SDK 0.4.12 poisti PrioritySettingsistä `fairness_weight`-kentän, mikä rikkoo kenttää käyttävän koodin.

08 Benchmarkit ja vertailut

Kaksi vertailujen tekijää vaihtoi viikolla mittatikkuaan. Artificial Analysis julkaisi keskiviikkona Intelligence Indexin menetelmäversion 5.0.1. Siinä CritPt-testin tekijöiden arviointi on tehty kestävämmäksi ja kaikki mallit on arvioitu uudelleen CritPt v2:lla. Epoch AI siirsi torstaina Epoch Capabilities Indexin bayesilaiseen malliin. Yhtiön mukaan se antaa paremmin perustellut epävarmuusarviot ja mahdollistaa mallien suoran vertailun. Kummassakaan tapauksessa vanhoja ja uusia lukuja ei voi rinnastaa tarkistamatta, onko historia laskettu uudelleen.

Viikon Intelligence Index -luvut jakautuvat menetelmämuutoksen molemmin puolin. Maanantaina Gemini 4 Carbon (Medium) sai 58 pistettä, DeepSeek V4.1 Flash 1017 (High) 29 ja Nanbeige4.3-33B 20 pistettä. Tiistaina Gemini 4 Carbonin Low-taso sai 54 pistettä. Perjantaina julkaistut tulokset ovat Muse Wonder (High) 41, Tencent Hy4 35 ja Inkling 2 Small 31 ja ne on laskettu uudella menetelmällä. Alkuviikon lukujen vertaaminen loppuviikon lukuihin edellyttää siis, että mallisivun nykyinen arvo tarkistetaan. Hinta- ja nopeuslukuja aineistossa ei ole, vaikka Flash-luokan mallin arvo ratkeaa yleensä juuri niissä.

Inkling 2 Small toi viikon kiinnostavimman koulutusluvun. Latent Spacen AINewsin mukaan 594 miljardin parametrin mallista on aktiivisena 17 miljardia. Malli koulutettiin 12 biljoonalla tokenilla 68 tunnissa. Karkea $6 \cdot N \cdot D$-laskelma aktiivisilla parametreilla antaa noin $1{,}2 \times 10^{24}$ laskutoimitusta. Se vastaa 68 tunnissa noin 5×10^{18} laskutoimitusta sekunnissa ja edellyttää tuhansien nykysukupolven kiihdyttimien klusteria. Lyhyt seinäkelloaika kertoo siis ennen kaikkea laskentakapasiteetista ja putken tehokkuudesta. Otsikon väite Muse Sparkin voittamisesta on julkaisijan oma ja vertailukohta jää epäselväksi. Artificial Analysis sijoittaa samassa erässä Muse Wonderin kymmenen pistettä Inklingin edelle. ARC Prize testasi mallin Latent Spacen mukaan. Tuloksia otteessa ei ole.

Mittareita rakentavat yhä useammin mallitoimittajat itse. OpenAI julkaisi viikon alussa LifeValin, joka mittaa tekoälyn apua arjen tehtävissä, kuten lastenhoidon etsimisessä, talousasioissa ja viranomaispapereissa. Lauantaina tuli BioBasicsBench: 100 tehtävää, joissa agentin on tulkittava kokeellista raakadataa ja hoidettava märkälaboratorion rutiineja. Kummankaan otteesta ei käy ilmi mitattuja malleja, tuloksia tai tehtäväsarjan julkisuutta. Biologia on samalla OpenAI:n Preparedness-kehikossa seurattu riskialue, minkä vuoksi laboratoriokyvyn mittari palvelee yhtä lailla riskiarviota kuin tuotekehitystä. Ennen kuin muut laboratoriot raportoivat tuloksiaan samoilla mittareilla, ne kuvaavat ensisijaisesti laatijansa tuotesuuntaa.

Mittaamisen pätevyyttä koetteli myös OpenAI:n alignment-tiimin tutkimus. Se tarkasteli, miten agenttien käytös eroaa turvallisuusarvioinneissa ja tuotannossa ja miten mallien tulkinnat arvioinnin tavoitteista liittyvät eroon. Testin tunnistava malli mittaa ensisijaisesti kykyään tunnistaa testejä. Samaan ongelmaan tarttui toisesta suunnasta Zou Groupin laaja toistotutkimus. Se ajaa vanhoja kielimalleista tehtyjä löydöksiä uusilla malleilla ja kysyy osoitteensa mukaan väitteiden säilyvyysaikaa. Hacker Newsissa se sai vain 30 pistettä. Ai2:n StateBench mittaa tilamuutosten seurantaa pitkissä dokumenteissa neljällä vaikeustasolla. Ai2:n mukaan kohdennettu koulutus parantaa erityisesti rekurrenttien ja hybridiarkkitehtuurien kykyä käsitellä pidempiä sekvenssejä. Hugging Facen blogissa julkaistu CWE-bench v2 laajentaa haavoittuvuusluokkien tunnistamisen vertailua. Open ASR Leaderboard lisäsi mittaukseen egyptiläisen ja nigerialaisen englannin.

Laskentatilastoissa Epoch AI raportoi Broadcomin TPU-toimitusten kaksinkertaistuneen toisesta kolmanteen neljännekseen. Toimitukset nousivat noin 800 000 siruun eli kahteen miljoonaan H100-ekvivalenttiin. Lauantaina julkaistu analyysi (Epoch AI 2026) arvioi Huawein Sisä-Mongolian Horinger-datakeskuksen Ascend-pohjaiseksi laskentakapasiteetiksi 7 % SpaceX:n Colossus 2:sta, jota Epoch pitää maailman suurimpana tekoälydatakeskuksena. Luku on arvio. Vertailu koskee kahta laitosta eikä maiden kokonaiskapasiteettia.

09 Työskentelytavat ja agentit

Agentti pysähtyy kysymään lupaa tuotantotietokantaan kirjoittamiseen. Hyväksyjä vastaa vartin päästä palattuaan kahvilta. Siinä välissä rajapinnan säie on vanhentunut viiden minuutin kohdalla ja tila, jota hyväksyntä koskee, on voinut muuttua. Moni viikon julkaisu osui juuri tähän väliin. LangGraph 1.2.20 antaa jokaiselle tehtävän `interrupt()`-kutsulle oman tunnisteen. Uusi tiukka jatkotila hylkää vanhentuneet vastaukset, mikä vähentää riskiä hyväksyä väärä toiminto ihmistä odottavassa ketjussa. SDK 0.4.14 toi saman `strict_resume`-valinnan kutsuihin `runs.create`, `stream` ja `wait`. Clinen yöversio delegoi käyttöliittymän verifioinnin aliagentille. Samalla delegoidut työkaluhyväksynnät, työntekijän identiteetti ja välitetty suoritustila säilyvät koko ketjun läpi. Datasetten laajennus datasette-apps 0.3a0 antaa agentille työkalut sovelluksen tallennettujen kyselyjen hallintaan. Kirjoituskyselyn myöntäminen vaatii silti käyttäjän hyväksynnän, jolloin agentin oikeus rajautuu yksittäiseen operaatioon.

Claude Codessa sama periaate eteni kahdessa vaiheessa. Tiistain 2.1.317 lisäsi Agent SDK:n hook-callbackeille asetuksen `onFailure: "block"`. Sen avulla epäonnistuva tarkistus estää toiminnon eikä päästä sitä läpi. Oletusarvoisen fail-openin rinnalle tuli näin fail-closed. Perjantain 2.1.319 muutti auto modea ympäristöissä, joissa telemetria tai feature flagit on kytketty pois, myös LLM-yhdyskäytävien takana. Kutsut, joita palvelinpuoli ei pysty tarkistamaan, joko kysytään käyttäjältä tai estetään. Paikalliseen sisäänrakennettuun luokittelijaan ei enää palata. Vanhan käytöksen saa takaisin asetuksella `CLAUDE_CODE_AUTO_MODE_SERVER=0`. Arvo `1` kytkee palvelintarkistuksen päälle myös silloin, kun `CLAUDE_CODE_DISABLE_EXPERIMENTAL_BETAS` on asetettu. Jälkimmäistä käyttävien hallinnoitujen asetusten on siksi asetettava myös `CLAUDE_CODE_AUTO_MODE_SERVER=1`. Bedrock, Vertex ja Foundry eivät muutu. Yhdyskäytävän takana työskentelevät tiimit näkevät todennäköisesti enemmän vahvistuskyselyjä, kunnes ylläpito valitsee linjansa. Viikonloppuna pysyvä lupa nousi otsikkoon toisesta syystä. Apiiron mukaan Claude Coden Don't Ask Again -toiminto oli tuonut salasanoja näkyviin 1 122 julkisessa GitHub-tiedostossa.

Yhden istunnon rinnalle työkalut rakentavat nyt usean agentin hallintaa. Codex CLI 0.167.0 avaa komennolla `codex agents --no-daemon` komentokeskuksen. Sekä komentokeskus että istuntovalitsin merkitsevät lukemattomat tulokset. Yhdistettyjen taustajärjestelmien välillä voi vaihtaa luonnosten säilyessä ja `gcp_setup_wizard` ohjaa Vertex AI:n käyttöönotossa Application Default Credentials -tunnuksilla. Gemini CLI alkoi kartoittaa SubagentStop-tapahtuman ja yhdistää koukkumäärittelyjä. Sen versio 0.67.0 tuo IDE-kumppanin konttihiekkalaatikoihin exec-stdion kautta, jolloin editori-integraatiota ei tarvitse uhrata eristyksen vuoksi. Claude Code 2.1.318:ssa modin `$.model.fork` hyväksyy `effort`-valinnan. Nopea sivutyö voi siis ajaa `low`-tasolla istunnon asetuksesta riippumatta. Valvomattomia ajoja varten `claude auth status` näyttää kirjautumisen päättymispäivän ja `claude -p` palauttaa JSONissa kentän `credential_expires_at`. Cursor lisäsi agenteilleen äänikeskustelun. Otsikosta ei selviä, vastaako agentti myös äänellä.

Agentin ja mallin välinen raja hämärtyi viikolla kahdesta suunnasta. Ollaman `ollama launch claude` soveltaa nyt mallin tulostetokenirajaa Claude Codelle, jolloin pilvimallit eivät enää törmää 128 000 tokenin oletuskattoon. Perjantaina `ollama launch` alkoi tukea myös Kimi Code CLI:tä. Claude Coden käyttö muiden mallien käyttöliittymänä on siis Ollaman tukema polku. Open WebUI 0.13.1 kääntää asetelman toisin päin. Open Terminalin tarjoamat koodausagentit, kuten Claude Code, Codex ja Pi, näkyvät mallivalitsimessa. Niiden kanssa voi keskustella kirjoittaen tai puhuen, valita päättelytason ja palvelutason ja hyväksyä työkalupyynnöt keskustelussa. Jokainen keskustelu käyttää yhtä agenttia. Agenttikeskustelua ei voi generoida uudelleen tai jatkaa, eikä viestiä voi muokata ja lähettää uudestaan. Edellinen versio 0.13.0 oli tuonut välilehtipohjaisen työtilan ja Desktop-välilehden, joka näyttää ja ohjaa kytketyn päätteen työpöytää. LLM-työkalun versio 0.37 lisäsi komennon `llm openai endpoint add` OpenAI-yhteensopivien päätepisteiden konfigurointiin. Lisäksi oletukseksi vaihtui gpt-6-luna, joka maksaa puolet aiemmasta oletuksesta.

Mallinvaihdon hinta näkyi Claude Coden omassa repossa. Triage-, dedupe- ja @claude-työnkulut siirrettiin claude-sonnet-5-5:lle. Triage oli siihen asti ajettu helmikuussa kiinnitetyllä claude-opus-4-6:lla. Ensimmäisen commitin perustelu, jonka mukaan CI-tilin API-pääsy hylkää Opus 5.5:n, tarkentui myöhemmin: asiaa ei ollut kokeiltu kyseisen repon tilillä. Sonnet 5.5:lle ei tarjota todo-työkaluja, minkä vuoksi komennosta poistettiin ohje "Make a todo list first". Kehote, joka olettaa tietyt työkalut, rikkoutuu mallinvaihdossa ilman virheilmoitusta.

Ajomoottori llama.cpp kirjasi koko viikon, millä työkaluilla sitä kehitetään. Maanantaina Assisted-by-rivit nimesivät Codexin muun muassa monen GPU:n all-reducen kilpatilanteiden korjauksissa, joissa virheet näkyvät epädeterministisenä tulosteena. Tiistaina ja keskiviikkona avustajina näkyivät Mac Studiolla (M5 Ultra, 256 Gt) llama.cpp:n kautta ajettu GLM-5.3-Flash ja M2 Ultralla (192 Gt) ajettu DeepSeek-V4-Flash-Vision-Exp. Perjantaina Codexin ja Clauden rinnalla oli paikallisesti ajettu Qwen3.8-Flash-Next laitteistotietoineen (RTX 5090 ja M5 Max 128 Gt). Sunnuntaina Codex esiintyi WebGPU:n COUNT_EQUAL-tuessa ja SpacemiT-suorittimien MXFP4-uudelleenpakkauksessa. Projekti kehittää moottoriaan osin malleilla, joita moottori itse ajaa. Avustus kirjataan Co-authored-by- ja Signed-off-by-merkintöjen rinnalle, jolloin avusteiset muutokset voi jäljittää katselmoinnissa ja regressioiden selvittelyssä. MCP-määrittelyn repossa agentit huomioitiin toisella tavalla. Perjantaina repoon lisättiin agenteille itsenäisesti luettava ohje PR-otsikoiden muodosta. Sunnuntaina dokumentoitu osallistumiskäytäntö edellyttää ennen pull requestin avaamista työryhmän hyväksyntää ja modelcontextprotocol-organisaation jäsenyyttä. Muut ohjataan issueihin ja työryhmiin.

Arvioinnissa viikko toi kolme erilaista vastausta siihen, mistä agentin virheet löydetään. Legoran teknologiajohtaja kuvasi Latent Spacessa oikeudellisen agentin kehittämistä tilanteessa, jossa data kuuluu asianajotoimistojen asiakkaille eikä yhtäkään tuotantojälkeä voi lukea. Evalit on silloin rakennettava ilman palautesilmukkaa, johon useimmat agenttitiimit nojaavat. NVIDIA kuvasi päinvastaisen mallin, jossa NeMo muodostaa tuotannon epäonnistumisista priorisoituja ongelmia. Niistä syntyy räätälöityjä arviointeja kehotteiden, työkalujen, muistin, taitojen ja harness-koodin parantamiseen. Microsoft Researchin PDD Kit muuntaa kehittäjän tarkoituksen Verus-määrittelyksi ja todistaa koodin täyttävän sen kaikilla syötteillä. Testeihin se turvautuu vain, kun todistus ei ole käytännöllinen. Lupaus tarkoituksen katselmoinnista rivien sijaan siirtää riskin määrittelyn oikeellisuuteen. Anthropicin perjantaina julkaisemasta engineering-kirjoituksesta *The components of agent memory* aineistossa on vain otsikko.

Vastaääni kuului samalla viikolla. The Vergen päätoimittajan mukaan agentit eivät vielä ole todellisia, eikä niitä käytetä niin kuin halutaan. Hacker Newsissa kehittäjän kirjoitus *How I use AI (Oct 2026)* keräsi 241 pistettä ja 280 kommenttia. Agenttityön käytännöt ovat siis yhä avoin keskustelu.

10 Liitännäiset, MCP ja repot

MCP:n viikon tärkein muutos ilmestyi sunnuntaina dokumentaation muodossa. Valtuutusopas kirjoitettiin uudelleen määrittelyversiolle 2026-07-28. Client ID Metadata Documents on nyt ensisijainen asiakkaan rekisteröintitapa ja Dynamic Client Registration on merkitty vanhentuvaksi. Toimittajakohtainen esimerkki korvattiin toimittajasta riippumattomilla vaatimuksilla. TypeScript-, Python- ja C#-esimerkit siirrettiin v2-SDK:ihin, joissa tarkistetaan yleisö, myöntäjä, laajuus ja alkuperä. Muutos koskee sekä 2026-07-28-versiota että luonnosta. Perjantaina Hacker Newsissa noussut kirjoitus käsitteli samaa ongelmaa käytännön tasolla. Sen mukaan palvelimen pitäisi antaa käyttäjän luottaa omaan MCP-asiakkaaseensa ilman, että uudelleenohjausosoitteiden hyväksyntä avaa reiän tietojenkalastelulle. Dynaamiseen rekisteröintiin nojaavat palvelintoteutukset ovat nyt vanhentuvalla polulla.

Aikataulu venyi samalla. Tiistaina blogikorjauksessa näkyi, että SEP-3398 siirsi seuraavan määrittelyjulkaisun joulukuulta 2026 maaliskuun loppuun 2027. Skillsin asema tarkentui sunnuntaina. Laajennus on SEP-3392:n mukaan beta-vaiheessa ja sen määrittely siirtyi hakemistosta `specification/stable/` hakemistoon `specification/beta/`. Tiistaina laajennusten tukimatriisiin oli lisätty Glean, joka tukee MCP Appsia, Enterprise-Managed Authorizationia, Skillsiä ja Tasksia. Claude Coden terminaali tukee versiosta 2.1.319 alkaen MCP-tehtäviä eli pitkäkestoisia työkalukutsuja, joita palvelin ajaa tehtävänä. Hacker Newsissa MCP Explained -selityssivusto keräsi 326 pistettä ja 133 kommenttia. Protokolla tavoittaa siis nyt myös lukijoita, jotka eivät ole sitä vielä käyttäneet.

Skills-muoto levisi viikolla myös MCP:n ulkopuolelle. Virallisessa Claude Code -liitännäiskokoelmassa Greptilen merkintä ohjattiin repon greptileai/skills versioon 1.3.1. GitHubin nousijoissa oli garden-skills, Claude Codelle, Cursorille ja Codexille tarkoitettu kokoelma, jossa on muun muassa web-suunnittelua, kuvagenerointikehotteita ja paikallinen tietohaku. Scenarion skills-kokoelma ohjaa koodausagentteja tuottamaan kuvaa, videota, ääntä ja 3D-sisältöä yhtiön MCP-palvelimen kautta. Käytännössä se rajautuu Scenarion käyttäjiin. Spotify jakaa Portal-liitännäisensä samasta reposta Claude Codelle, Codexille,

Cursorille ja pi:lle. Liitännäinen tuo agenttiin tunnistautumisen, ohjelmistokatalogin haun, palvelukatsaukset ja vain luku -diagnostiikan. Portalin käyttäjille se on kokeilun arvoinen ja muille se on malli saman liitännäisen jakelusta usealle agentille. Kolmannen osapuolen skill ajaa koodia agentin oikeuksin, minkä vuoksi se on luettava ennen asennusta samalla tavalla kuin mikä tahansa ajettava riippuvuus.

Vercel Labsin kaksi viikon trendirepoa rakentavat agentille saman katseen eri pinnoille. Tiistaina noussut agent-desktop on komentorivityökalu ja C-kirjasto computer useen macOS:ssä, Windowsissa ja Linuxissa. Se toimii käyttöjärjestelmän saavutettavuuspuun kautta. Snapshot palauttaa JSONin, jossa jokaisella toiminnon kohteella on viite, jolloin pikselivertailua ei tarvita. Jokainen kutsu ajetaan omassa prosessissaan ja headless-tilassa etualalle nosto ja synteettinen syöte on estetty. Perjantaina noussut agent-browser tekee saman selaimelle. Rust-pohjainen työkalu pitää selaimen auki komentojen välillä ja palauttaa sivun tiiviinä saavutettavuuspuuna, jossa elementeillä on viitteet kuten `@e1`. Agentti toimii viitteen perusteella, lukee sivun uudelleen ja tarkistaa tuloksen, jolloin jokainen askel on yksi lyhyt shell-kutsu raskaan selainkehyksen sijaan. Molemmat ovat kokeilun arvoisia, koska saavutettavuuspuu on kuvakaappauksiin verrattuna deterministisempi rajapinta. Sama rakenne on alun perin tehty ihmisille, joille ruudun kuva ei kerro tarpeeksi.

Jäljitettävyys oli toinen repojen teema. AiSOC on itse ylläpidettävä avoimen lähdekoodin tietoturvavalvomo, joka luokittelee hälytykset LLM-agenteilla ja tutkii tapaukset MITRE ATT&CK -kehyksen mukaan. Jokainen agentin askel tallentuu toistettavaan päätöslokiin. Valvomo toimii ilman API-avaimia, on MIT-lisensoitu ja tuo MCP-palvelimen Claudelle, Cursorille ja Continuelle. Laboratorioympäristössä se on kokeilun arvoinen juuri auditoitavan lokinsa vuoksi. Tuotantokäytöstä ei ole riippumatonta näyttöä. Valvomon ja koodieditorin välinen MCP-yhteys on lisäksi itsessään uusi luottamusraja. Torstain Show HN -nosto Stroq lupaa paikallisen kirjanpidon siitä, mitä koodausagentti teki ja miksi. Se keräsi 93 pistettä ja 34 kommenttia, mutta aineistossa siitä on vain otsikko.

Sunnuntaina noussut billion-context on kontekstin pakkausliitännäinen. Se lupaa kuukauden mittaisia yksittäisiä istuntoja, viisinkertaisen token-säästön ja sen, että 100 000 tokenin ikkuna riittää. Terveen istunnon etuliitevälimuistin osumaosuudeksi repo ilmoittaa 95–97 %. Luvut ovat tekijän omia. Pakkauksen laatuväite ratkeaa vasta omilla istunnoilla, koska pakattu konteksti on aina alkuperäistä suppeampi. Metan astryx on avoimeksi julkaistu design system. Se on kasvanut yhtiön sisällä kahdeksan vuotta ja palvelee yhtiön mukaan yli 13 000 sovellusta. Järjestelmä on beta-vaiheessa, rakentuu React 19:n ja StyleX:n varaan ja on kuvauksen mukaan agenttivalmis. Mitä agenttivalmius tarkoittaa, otteesta ei selviä.

Kehitysympäristöön sopii floci, ilmainen paikallinen AWS-emulaattori. Siihen SDK, CLI, Terraform, CDK ja testit ohjataan osoitteeseen `localhost:4566`. Se on kokeilun arvoinen kehityksessä ja CI:ssä, kunhan palveluiden kattavuus tarkistetaan omaa käyttöä vasten. NVIDIA/auto-ontology rakentaa datan päälle semanttisen graafikerroksen deterministisiä kyselyjä varten. Se sopii analytiikka-agentteihin, joissa väärä luku on kalliimpi kuin hidas vastaus. The-PR-Agent/pr-agent on Qodon yhteisön ylläpitämä perintöprojekti eikä Qodon ilmaisversio. Ero vaikuttaa tuen ja jatkuvuuden arviointiin. Paikallisista sanelusovelluksista Handy ja hex lataavat puhemallin kerran ja tekevät tunnistuksen sen jälkeen laitteella ja kumpikin on kokeilun arvoinen.

Kohinaa oli runsaasti. Kokoelman awesome-claude-code commitit olivat koko viikon automaattisia tikkeripäivityksiä. Virallisen liitännäiskokoelman kumppanipäivitykset olivat enimmäkseen versionostoja ilman toiminnallisia kuvauksia. Viikonlopun trendilistan kärjessä oli vakiintuneita projekteja, joilla ei ollut tekoälyuutista.

11 Tietoturva ja haavoittuvuudet

Viikon puhutuimmat tietoturvatapaukset eivät koskeneet malleja vaan agenttityökalujen ympäristöä: lisäosarekisteriä, verkkopyyntöjen otsakkeita, tallennettuja lupia ja tilin asetuksia. Kolme suurinta keskustelua liittyi Claude Codeen ja niiden pääketjut keräsivät Hacker Newsissa yhteensä yli 3 000 pistettä.

Maanantaina Phorionin kirjoitus (1 006 pistettä, 553 kommenttia) kuvasi kompromisoidun julkaisun Claude Coden virallisesta lisäosasta Open VSX -rekisterissä. Cursor ja muut VS Code -haarat käyttävät tätä rekisteriä Microsoftin markkinapaikan sijaan. Cursor julkaisi asiasta tietoturvailmoituksen. DoublePulsarin mukaan haittakuorma oli etäkäyttötroijalainen. Kirjoituksen otsikko esittää

Anthropicin tulleen murretuksi. Tietoturvatutkijan Mastodon-viesti puhuu monikossa takaovista virallisissa lisäosissa. Julkisista otsikoista ei voi todentaa, oliko murrettu julkaisutili, rakennusputki vai rekisteri. Myöskään sitä, koskeeko tapaus Microsoftin markkinapaikalta asennettuja versioita, ei voi otsikoista päätellä. Haavoittuneet versiot ja julkaisuikkuna on yhä tarkistettava Cursorin ilmoituksesta ja Phorionin kirjoituksesta, koska viikon myöhemmät lähteet eivät tuoneet niihin lisätietoa.

Tunnistus alkaa editorin lisäosaluettelosta. Lisäosan tunnistetta, versiota ja asennusajankohtaa verrataan ilmoitettuun ikkunaan. Lisäksi tarkistetaan lisäosahakemisto, joka on Cursorissa oletuksena `~/.cursor/extensions`. Kone käsitellään murrettuna, jos saastunut versio on ollut asennettuna. Pelkkä poisto ei riitä. Palveluiden API-avaimet, GitHub-tokenit, SSH-avaimet, pilvitunnukset ja `.env`-tiedostojen salaisuudet uusitaan. Pysyvyysmekanismit, kuten launchd-agentit, cron-ajot ja shellin alustustiedostot, käydään läpi. Kriittisissä ympäristöissä lisäosien automaattinen päivitys kytketään pois ja versiot kiinnitetään.

Sähköpostiosoitteen tapaus eteni viikon aikana kahdessa vaiheessa. Tiistain Claude Code 2.1.317 alkoi kysyä lupaa ennen kuin verkko- tai shell-työkalukutsu lähettää tilin sähköpostiosoitteen verkkoon. Perjantaina issue #107757 otsikolla "Claude Code sends your email address to websites" keräsi Hacker Newsissa 1 512 pistettä ja 475 kommenttia. Vastauksena repoon yhdistettiin muutos #107768, joka julkaisee sisäänrakennetun privacy-modin lähdekoodin. Modi lukee tilin tunnisteen uudesta `$.session.identity`-kentästä. Se hylkää shell-kutsun, joka sisältää kirjautuneen tilin sähköpostiosoitteen ja nimeää User-Agent-otsakkeen. Toinen commit dokumentoi asetukset, jotka kytkevät modin pois, minkä vuoksi muokatuilla asetuksilla työskentelevän on tarkistettava ne. Versiota, jossa suoja on käytössä, aineistossa ei kerrota. Altistumisen voi tunnistaa etsimällä agentin istuntolokeista ja shell-historiasta komentoja, joissa oma osoite esiintyy User-Agent-arvona, esimerkiksi muodoissa `curl -A` tai `-H "User-Agent: ..."`.

Lauantai-iltana Hacker Newsiin nousi Apiiron kirjoitus otsikolla "Claude Code's 'Don't Ask Again' exposes passwords in 1,122 public GitHub files". Se keräsi vuorokaudessa 514 pistettä ja 375 kommenttia. Mekanismi on tarkistettava Apiiron kirjoituksesta, koska aineistossa on vain otsikko. Ennen commitia on aiheellista käydä läpi agentin projektihakemistoon kirjoittamat asetus- ja lupatiedostot ja ajaa repoon salaisuusskanneri. Salasana, joka on ollut agentin hyväksymässä komennossa, vaihdetaan, jos tiedosto on päätynyt julkiseen repoon. Sunnuntaina The State of AI -sivuston uutisen mukaan Goldman Sachs vetää Claude Coden pankkiiriensa ja kaupankävijöidensä käytöstä. Syy ei käy otteesta ilmi. Uutinen keräsi 31 pistettä ja kaksi kommenttia. OpenAI:n puolella tiistaina raportoitiin, että ChatGPT:n data-asetusten kohta "Improve the model for everyone" oli palannut Codex-keskusteluissa päälle käyttäjän sitä valitsematta. Hacker Newsin ketju viittasi aiempaan samanlaiseen havaintoon. Organisaatioissa asetuksen tilaa ei voi pitää pysyvänä kontrollina.

Ajomoottoreista llama.cpp korjasi viikon aikana kaksi muistin rajojen ylitystä, joita julkaisumerkinnät eivät luokittele haavoittuvuuksiksi. Tiistain b12073 korjasi Vulkan-taustan rajojen ylittävän luvun NVFP4:n UE4M3-skaalojen haussa. Ei-OCP-polulla `ue4m3_fp32_lut`-taulukossa oli 128 merkintää ja sitä indeksoitiin raa'alla skaalatavulla. `ggml_validate_row_data` hyväksyy minkä tahansa tavun ja `ggml_ue4m3_to_fp32` ohittaa bitin 7, jolloin tavu 0x80 tai suurempi lukee jaetun taulukon ohi. Vulkanin robustness-ominaisuus ei kata jaettua muistia. Korjaus kasvattaa taulukon, alustussilmukat ja isännän jaetun muistin budjetit 256 merkintään. Altis on Vulkan-taustaa NVFP4-malleilla käyttävä build ennen b12073:a.

Keskiviikon b12102 korjasi ggml:n näkymien rajatarkistuksen, joka oli käyttänyt saman muotoisen yhtenäisen tensorin kokoa ennen strideja. Leveämmillä strideillä näkymä saattoi läpäistä tarkistuksen ja silti ulottua lähdepuskurin ohi. Nyt konstruktori tarkistaa todellisuudessa saavutettavat tavut. Lisäksi b12078 hylkää GGUF-jakotiedostojen yhdistämisessä yhteensopimattomat osat. CVE-tunnisteita aineistossa ei ole. Riski koskee erityisesti epäluotettavista lähteistä ladattuja GGUF-tiedostoja, joita ei ole syytä ajaa b12102:ta vanhemmilla buildeilla. Sama varovaisuus koskee trendilistan riisuttuja johdannaisia. Alkuperä varmistetaan julkaisijan organisaatiotililtä. Tiedostomuotona safetensors on turvallisempi kuin pickle-pohjaiset formaatit, koska se ei suorita koodia latauksen yhteydessä.

MCP:n referenssipalvelimien 4.11. julkaisukierros korjasi kaksi riippuvuusadvisorya. Filesystem-palvelin nosti riippuvuuden `minimatch ^11.0.4`:ään advisoryn GHSA-g3hv-75wg-m249 vuoksi. Commitin mukaan kaikki 10.x-versiot olivat korjatun 11.0.1:n alapuolella. Riippuvainen paketti ei voinut korjata asiaa käyttäjilleen overridella, koska npm lukee overridet vain juuriprojektista. Oman puun tilanteen näkee komennolla `npm ls minimatch`. Toinen

korjaus nosti `qs`-overriden tasolle `>=6.16.1` advisoryn GHSA-6pq8-jhm2-2vcw vuoksi. Sen tilan näkee komennolla `npm ls qs`. Sequentialthinking-palvelimen stderr-laatikko täytti jokaisen rivin pisimmän rivin mittaiseksi, jolloin 60 kilotavun ajatus saattoi tuottaa 400 miljoonaa merkkiä lokiin. Laatikko rajataan nyt 100 merkin leveyteen ja 1 000 riviin ja koko jää noin 107 kilotavuun. Korjaus on tehty 4.11. versioleimauksen jälkeen, joten sitä ei pidä olettaa kyseisen päivän paketteihin. Tietoturvaohjeisiin lisättiin käytäntö viedä korjaus v2/mainiin läpinäkymättömän julkisen PR:n kautta.

Agenttien polkurajoja koskivat Gemini CLI:n muutokset. Versio 0.68.0-preview.0 tekee `isWithinRoot`-tarkistuksesta kirjainkoosta riippumattoman Windowsissa. Versio 0.67.0 rajasi luottamattoman tilan lipun tarkistusta. Kumpaakaan ei ole merkitty haavoittuvuudeksi, vaikka polkurajat ovat agenttien eristyksen ydintä. Anthropicin virallisen kokoelman security-guidance-liitännäinen sai perjantaina toistakymmentä kovennusta epäluotettavaa repoa vastaan. Windowsilla git (#7004) ja Agent SDK:n CLI (#6999) käynnistetään täydellä polulla, mikä estää repon hakemistoon istutettua ohjelmaa ohittamasta oikeaa. Projektin asetustiedostot luetaan vain oikeina tiedostoina (#7003), mikä torjuu symbolisilla linkeillä tehdyt temput. Git log luetaan ilman allekirjoitustarkistuksia (#7002), jolloin repon asetukset eivät pääse käynnistämään allekirjoitusohjelmaa. Tilakansio rajataan omistajalleen (#7001) ja katselmointiajoja ei enää tallenneta istuntotranskripteiksi (#6627). Lauantaina liitännäinen nousi versioon 2.0.18. Samalla tuli luotettavuuskorjauksia, joiden jälkeen aikakatkaistu SDK-asennus ei jätä keskeneräistä latausta ja indeksin kopiosta pudotetaan alimoduulien gitlink-merkinnät. LangGraph CLI 0.4.42 päivitti filelock- ja langgraph-api-riippuvuudet tietoturvakorjausten vuoksi nimeämättä tunnisteita.

Tivin raportoimassa tapauksessa kaapattu yrityksen tekoälytili tuotti 100 000 dollarin laskun minuuteissa. Kaappaustapaa uutisen otteessa ei kerrota. Palvelukohtainen IP-rajaus, avainten kierto ja kulutuspoikkeamien hälytykset ovat tunnettuja vastatoimia. Ne rajaavat vahinkoa kuitenkin vain, jos ne on otettu käyttöön ennen vuotoa.

Kyberturvallisuuskeskuksen viikon lanka koski fyysisiä laitteita. Tiistain tiedotteen mukaan Euroopassa lokakuussa alkanut hyökkäysaalto kohdistui kauppojen kylmäjärjestelmien etäohjattaviin laitteisiin. Hyökkääjät muuttivat lämpötila- ja sulatusasetuksia. Marraskuun alussa vastaavia hyökkäyksiä raportoitiin useissa maissa, myös Virossa. Epäiltyjä tapauksia havaittiin myös Suomessa. Perjantain viikkokatsauksen 45/2026 mukaan kylmälaitteiden ilkivaltainen hyväksikäyttö on aiheuttanut vahinkoja Suomessakin. Tivin mukaan Ranskassa pakastimiin iskenyt aalto taas torjuttiin Suomessa varoitusten ja varautumisen ansiosta. Kumpikaan lähde ei nimeä laitevalmistajaa eikä haavoittuvuustunnistetta. Tunnistaminen alkaa inventaariosta, josta selviää, mitkä säätölaitteet näkyvät julkiseen verkkoon ja millä tunnuksilla niihin pääsee. Seuraavaksi hallintaliittymät siirretään VPN:n tai erillisen verkon taakse ja oletustunnukset vaihdetaan. Huoltoyhtiöiden etäportaaleihin otetaan monivaiheinen tunnistautuminen ja laiteohjelmistot päivitetään.

Lähteet

A thousand fewer paths to root · OpenAI

GPT-6 Cyber: Advancing the frontier of aligned cybersecurity agents · OpenAI

Partnering with Hacktron to find vulnerabilities in OpenAI's systems · OpenAI

Rewarding defenders with bigger bug bounties · OpenAI

Preparing our next security sprint to harden OpenAI · OpenAI

Daybreak Red is now available in AWS GovCloud (US-East) · OpenAI

Scaling Git at OpenAI · OpenAI

OpenAI to acquire Pierre Computer Company, Inc. · OpenAI

Introducing Images 2.6 · OpenAI

Training activation classifiers for safety and alignment · OpenAI Alignment Research Blog

Eval-production behavior gaps in agent safety and the role of eval-awareness · OpenAI Alignment Research Blog

LifeVal · OpenAI

Measuring how AI agents handle everyday wet-lab science · OpenAI

Claude Platform release notes, November 2, 2026 · Anthropic API release notes

Claude Platform release notes, November 3, 2026 · Anthropic API release notes

Claude Platform release notes, November 5, 2026: Admin API sign out · Anthropic API release notes

Claude Platform release notes, November 5, 2026: IP allowlisting · Anthropic API release notes

Progress on provable inference for large language models · Hacker News: MCP

The components of agent memory · Anthropic engineering

Independent Assessment of AI R&D Acceleration at Anthropic in October 2026 · METR

Review of the "Risks from automated R&D" section in Anthropic's Risk Report (August 2026) · METR

[AINews] Anthropic IPO: Whither Now? · Latent Space

Gemini 4 Carbon Heads Toward Wider Availability, What Anthropic's IPO Means for Developers, Genie 4 Combines with Gemini to Train Robots · The Batch (deeplearning.ai)

Speeding up planning decisions with AI · Google DeepMind

Agentic cheminformatics with Google Antigravity · Google DeepMind

Gemini Omni 1.2 Flash: Fast and cost-effective with longer clips · Google DeepMind

Le Chonk brings the brains. Studio puts it to work. · Mistral

Giving every team a data scientist with Grok Bot · xAI news

LongCat-2.5: Native Multimodality for Agentic Tasks · Hacker News: agents

We're training a Figma-fluent editing model · Hacker News: MCP

China's First Frontier AI Export Controls: Mofcom Announcement No. 48 · Hacker News: agents

[AINews] LLMs can read & write neural network weights · Latent Space

Qwen 3.8 27B on my laptop can interpret model weights, too · Simon Willison

p-e-w/heretic · GitHub trending -syöte

Trending: AtomicChat/DeepSeek-V4.1-Flash-1017-Uncensored-GSQ-RCO-GGUF · HF trending -mallit

OBLITERATUS/DeepSeek-V4.1-Flash-1017-obliterated · HF trending -mallit

xtools-at/DeepSeek-V4-Pro-Qwen3.5-9B-UNCENSORED-heretic · HF trending -mallit

ccore/Qwen3.5-4B-GLB-3D-Coder · HF trending -mallit

Trending: moonshotai/Kimi-K3.1-Pro · HF trending -mallit

[AINews] It's Swapping Season · Latent Space

Trending: ricursiveAI/Ripple-65B-Preview · HF trending -mallit

Ripple-65B Preview: U.S. Open-Weight SOTA on Agentic Coding · Hacker News: agents

Trending: NousResearch/Arete-27B · HF trending -mallit

How Olmo 3 earned the highest ethical rating from Nextcloud · Ai2 Blog

Show HN: A decent text-to-image model trained for ~$20K (Apache 2.0 Weights) · Hacker News: MCP

Show HN: We compressed a 31B video+audio diffusion model to 14B, 6 steps, INT8 · Hacker News: MCP

LiquidAI/LFM3-Omni-2B · HF trending -mallit

New NVIDIA Cosmos 3 Open Models for Interactive World Simulation · Hugging Face -blogi

ToyBlocks: A controlled world for robot learning and evaluation · Ai2 Blog

Koodarien tekoälykulut karkasivat käsistä, Revolut pudotti laskunsa murto-osaan yhdellä ratkaisulla · Tivi

Closed AI models dominate the price-performance frontier · Hacker News: MCP

How big is the open model economy today? · Interconnects

FareedKhan-dev/kimi-k3-in-c · GitHub trending -syöte

llama.cpp b12170 · llama.cpp-julkaisut

b12191: llama_moe_cache: run missed experts on CPU, add tuning script (#30966) · llama.cpp-julkaisut

b12206 · llama.cpp-julkaisut

FlashML-org/FreeToken · GitHub trending -syöte

Qualcomm Cloud AI 100 Ultra, 150W AI Accelerator Card with 128GB at 548GB/s · Hacker News: agents

b12051: vulkan: add tensor-parallel allreduce function · llama.cpp-julkaisut

b12048 · llama.cpp-julkaisut

llama.cpp b12079: Add nccl support for multi-node tensor parallelism · llama.cpp-julkaisut

v0.9.0 · llama.cpp-julkaisut

b12100 · llama.cpp-julkaisut

b12207 · llama.cpp-julkaisut

b12127: metal: keep fusable node groups inside one command buffer (#31688) · llama.cpp-julkaisut

b12189 · llama.cpp-julkaisut

b12208 · llama.cpp-julkaisut

b12203: ggml : switch to F32 src1 in the CPU backend (#31854) · llama.cpp-julkaisut

b12205: webgpu : add COUNT_EQUAL support (#31578) · llama.cpp-julkaisut

b12210 · llama.cpp-julkaisut

b12182 · llama.cpp-julkaisut

b12183 · llama.cpp-julkaisut

llama.cpp b12169: ci : bump CUDA 13 release builds to 13.5 · llama.cpp-julkaisut

Ollama v0.40.14 · Ollama-julkaisut

Ollama v0.40.15 · Ollama-julkaisut

Don't Call It a Decision Model · Hacker News: MCP

Making Kubernetes AI-Ready for NVIDIA Vera Rubin on Day One · NVIDIA Technical Blog

NVIDIA CUDA 13.5 Brings NVIDIA Vera Rubin GA Support · NVIDIA Technical Blog

dsx-ai-factory/nvke · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

Accelerating Reinforcement Learning Weight Refit with the Reshard Collective · NVIDIA Technical Blog

Launch HN: Rivet (YC W23), Scale and Orchestrate Agents on S3 · Hacker News: agents

n8n@2.47.0 · n8n releases

n8n@3.1.0 · n8n releases

n8n@2.46.4 · n8n releases

langgraph-sdk==0.4.12 · LangGraph releases

Artificial Analysis Intelligence Index methodology · Artificial Analysis

Updating the Epoch Capabilities Index methodology to use a Bayesian model · Epoch AI

Gemini 4 Carbon (Medium) · Artificial Analysis

Gemini 4 Carbon (Low), Intelligence Index: 54 · Artificial Analysis

DeepSeek V4.1 Flash 1017 (High) · Artificial Analysis

Nanbeige4.3-33B · Artificial Analysis

Muse Wonder (High), Intelligence Index: 41 · Artificial Analysis

Tencent Hy4, Intelligence Index: 35 · Artificial Analysis

Inkling 2 Small, Intelligence Index: 31 · Artificial Analysis

Inkling 2 Small: 17A-594B trained in 68 hours on 12T tokens, beating Muse Spark · Latent Space

Mass attempted replication of old findings of LLMs with new models · Hacker News: MCP

StateBench: Testing how language models track state changes over long documents · Ai2 Blog

CWE-bench v2: More coverage, quality, and transparency · Hugging Face -blogi

Egyptian and Nigerian English join the Open ASR Leaderboard · Hugging Face -blogi

Broadcom TPU compute shipments doubled from Q2 to Q3 2026, reaching 2 million H100-equivalents · Epoch AI

What Huawei's Horinger data center reveals about China's compute gap · Epoch AI

langgraph==1.2.20 · LangGraph releases

langgraph-sdk==0.4.14 · LangGraph releases

desktop-nightly-20261104233655: feat(core): delegate UX verification to subagents (#15523) · Cline releases

datasette-apps 0.3a0 · Simon Willison

Claude Code 2.1.317 · Claude Code changelog

Claude Code 2.1.318 · Claude Code changelog

Claude Code 2.1.319 · Claude Code changelog

OpenAI Codex CLI 0.167.0 · OpenAI Codex CLI releases

Release v0.67.0-nightly.20261104.g89e9d25d6 · Gemini CLI releases

Release v0.67.0 · Gemini CLI releases

Voice Chat with Agents · Cursor blog

Open WebUI v0.13.0 · Open WebUI releases

v0.13.1 · Open WebUI releases

llm 0.37 · Simon Willison

Move issue triage and dedupe workflows to Claude Sonnet 5.5 · Claude Code repo commits

Move @claude mentions to Claude Sonnet 5.5 · Claude Code repo commits

docs: make agent PR title guidance self-contained · MCP-määrittely commits

Document working-group-based contribution policy (#3514) · MCP-määrittely commits

Building Legal Agents When You Can't See the Data · Latent Space

Use NVIDIA NeMo to Turn Agent Failures into Tests and Harness Fixes · NVIDIA Technical Blog

Formally verifying agent-written full-stack web apps with PDD Kit · Microsoft Research

Quoting Nilay Patel · Simon Willison

How I use AI (Oct 2026) · Hacker News: agents

docs: rework the authorization guide for the 2026-07-28 specification… (#3518) · MCP-määrittely commits

Letting users trust their own MCP clients, without opening a phishing hole · Hacker News: MCP

Update the Ruby SDK Tier 1 blog post (#3496) · MCP-määrittely commits

Point Skills specification links at specification/beta/ (#3466) · MCP-määrittely commits

Add Glean to extension support matrix (#3495) · MCP-määrittely commits

MCP Explained · Hacker News: MCP

repoint(greptile): greptileai/claude-plugin → greptileai/skills @ v1.3.1 · Claude Code plugins/marketplace (virallinen)

ConardLi/garden-skills · GitHub trending -syöte

scenario-labs/skills · GitHub trending -syöte

spotify/portal-ai-plugins · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

vercel-labs/agent-desktop · GitHub trending -syöte

vercel-labs/agent-browser · GitHub trending -syöte

beenuar/AiSOC · GitHub trending -syöte

Show HN: Stroq, a local audit of what your coding agent did and why · Hacker News: agents

ranxianglei/billion-context · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

meta/astryx · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

floci-io/floci · GitHub trending -syöte (mshibanami/GitHubTrendingRSS)

NVIDIA/auto-ontology · GitHub trending -syöte

The-PR-Agent/pr-agent · GitHub trending -syöte

anomalyco/hex · GitHub trending -syöte

cjpais/Handy · GitHub trending -syöte

chore: update repo ticker data and SVGs [skip ci] · awesome-claude-code commits

Backdoor in Claude Code for VS Code extension on the Cursor Marketplace · Hacker News: Claude Code

Claude Code extension security notice · Cursor (Hacker News -ketjun kautta)

Claude Code VS Code extension with backdoor released on Cursor marketplace · Hacker News: Claude Code

Backdoors in official Claude Code extensions · Hacker News: Claude Code

Claude Code sends your email address to websites · Hacker News: Claude Code

mods: add the privacy mod (#107768) · Claude Code repo commits

Claude Code's "Don't Ask Again" exposes passwords in 1,122 public GitHub files · Hacker News: Claude Code

Goldman Sachs Is Pulling Claude Code from Its Bankers and Traders · Hacker News: Claude Code

OpenAI seems to be silently re-enabling model training on Codex chats · Hacker News: MCP

llama.cpp b12073 · llama.cpp-julkaisut

b12102 · llama.cpp-julkaisut

llama.cpp b12078 · llama.cpp-julkaisut

fix(filesystem): move minimatch to ^11.0.4 for GHSA-g3hv-75wg-m249 · MCP servers commits

fix(deps): raise the qs override to >=6.16.1 for GHSA-6pq8-jhm2-2vcw · MCP servers commits

fix(sequentialthinking): bound the stderr thought box · MCP servers commits

chore: stamp Python CalVer versions for 2026.11.4 · MCP servers commits

docs(security-advisory): pass --head when opening the public PR from … · MCP servers commits

Release v0.68.0-preview.0 · Gemini CLI releases

security-guidance: start git by its full path on Windows · Claude Code plugins/marketplace (virallinen)

security-guidance: on Windows, name the CLI for the Agent SDK by its full path · Claude Code plugins/marketplace (virallinen)

security-guidance: read project config files only as real files · Claude Code plugins/marketplace (virallinen)

security-guidance: read git log output without signature checks · Claude Code plugins/marketplace (virallinen)

security-guidance: keep the state folder to its user · Claude Code plugins/marketplace (virallinen)

security-guidance: do not save review runs as session transcripts · Claude Code plugins/marketplace (virallinen)

security-guidance: a timed-out SDK install leaves no download behind and pauses before retrying (#6985) · Claude Code plugins/marketplace (virallinen)

security-guidance: drop every gitlink from the index copy (#7000) · Claude Code plugins/marketplace (virallinen)

security-guidance: bump version to 2.0.18 (#7014) · Claude Code plugins/marketplace (virallinen)

langgraph-cli==0.4.42 · LangGraph releases

Hakkerit ottivat yhtiön tekoälytilin käyttöön ja raksuttivat minuuteissa 100 000 dollarin laskun · Tivi

IoT- ja OT-ympäristöjen tietoturva alkaa laitteiden suojaamisesta · Kyberturvallisuuskeskus

Kyberturvallisuuskeskuksen viikkokatsaus, 45/2026 · Kyberturvallisuuskeskus

Asiantuntija arvioi: Kauppojen kylmälaitteisiin kohdistunut hyökkäysaalto "on vasta alkua" · Tivi

Pakastimiin Ranskassa iskenyt kyberhyökkäysaalto onnistuttiin torjumaan Suomessa, Varautuminen kannatti · Tivi